A TRIDENT SCHOLAR PROJECT REPORT

NO. 457

Path Planning for Reduced Identifiability of Unmanned Surface Vehicles Conducting Intelligence, Surveillance, and Reconnaissance

by

Midshipman 1/C Ethan W. Doherty, USN

UNITED STATES NAVAL ACADEMY
ANNAPOLIS, MARYLAND

USNA-1531-2

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
05-22-17		

4. TITLE AND SUBTITLE

Path Planning for Reduced Identifiability of Unmanned Surface Vehicles Conducting Intelligence, Surveillance, and Reconnaissance

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Doherty, Ethan W.

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Naval Academy
Annapolis, MD 21402

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

Trident Scholar Report no. 457 (2017)

12. DISTRIBUTION / AVAILABILITY STATEMENT

This document has been approved for public release; its distribution is UNLIMITED.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Unmanned Surface Vehicles (USVs) provide the Navy with the capability to perform many dull, dirty, and dangerous operations. Intelligence, Surveillance, and Reconnaissance (ISR) in hostile areas has been identified as one of the mission areas where USVs can be particularly effective. These missions involve having a vessel enter enemy waters in order to gather information about the adversary's coastal waterways, naval vessels, and maritime installations. The objective of this research is to plan paths within hostile waters that a USV can follow in order to conduct ISR missions without being identified as a suspicious vessel. This involves the two competing objectives of gathering high quality intelligence and avoiding identification by the adversary.

This research focuses on the pre-mission planning of the path that a USV should follow in order to conduct an ISR mission. The paths are evaluated in order to determine the quality of the intelligence gathered by the USV and its risk of identification during its mission. This path planning process is tested on a number of sample missions in order to verify its suitability at developing ISR mission paths for USVs operating in a range of geographic areas.

15. SUBJECT TERMS

ship routing, path planning, optimal control, hypothesis test

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**		80	**19b. TELEPHONE NUMBER** *(include area code)*

U.S.N.A. --- Trident Scholar project report; no. 457 (2017)

PATH PLANNING FOR REDUCED IDENTIFIABILITY OF UNMANNED SURFACE VEHICLES CONDUCTING INTELLIGENCE, SURVEILLANCE, AND RECONNAISSANCE

by

Midshipman 1/C Ethan W. Doherty
United States Naval Academy
Annapolis, Maryland

_____
(signature)

Certification of Advisers Approval

Professor Kiriakos Kiriakidis
Systems Engineering Department

_____
(signature)

_____
(date)

CAPT Michael A. Hurni, USN
Systems Engineering Department

_____
(signature)

_____
(date)

Acceptance for the Trident Scholar Committee

Professor Maria J. Schroeder
Associate Director of Midshipman Research

_____
(signature)

_____
(date)

USNA-1531-2

Abstract

Unmanned Surface Vehicles (USVs) provide the Navy with the capability to perform many dull, dirty, and dangerous operations. Intelligence, Surveillance, and Reconnaissance (ISR) in hostile areas has been identified as one of the mission areas where USVs can be particularly effective. These missions involve having a vessel enter enemy waters in order to gather information about the adversary's coastal waterways, naval vessels, and maritime installations. The objective of this research is to plan paths within hostile waters that a USV can follow in order to conduct ISR missions without being identified as a suspicious vessel. This involves the two competing objectives of gathering high quality intelligence and avoiding identification by the adversary.

This research focuses on the pre-mission planning of the path that a USV should follow in order to conduct an ISR mission. The USV will attempt to blend into the normal maritime traffic by mimicking the behavior of the vessels within its area of operations. Automatic Identification System (AIS) transmissions from ships in the area of interest are used to determine the normal traffic behavior. This behavior is then integrated into the framework of an optimal control problem in order to solve for the USV's ideal path. Within this framework, a cost function models the USV's mission objectives of collecting intelligence and avoiding identification. DIDO software then solves the optimal control problem for candidate solutions. These potential solution paths are first validated by checking a number of their mathematical characteristics. The paths are then evaluated in order to determine the quality of the intelligence gathered by the USV and its risk of identification during its mission. This path planning process is tested on a number of sample missions in order to verify its suitability at developing ISR mission paths for USVs operating in a range of geographic areas.

Keywords:

ship routing, path planning, optimal control, hypothesis test

Acknowledgments

I would like to thank Professor Kiriakidis and CAPT Hurni for their guidance and instruction throughout the entire research process. The success of this research and my development as a student are thanks to their patient, repeated explanations and discussions. I also wish to thank both of them for their flexibility and support following my accident and throughout my recovery.

I also wish to thank the Trident Scholar Committee and Professor Schroeder for the opportunity to participate in this defining project of my academic career. Their trust in me and understanding of my situation following my accident are also greatly appreciated. I thank my academic adviser, Dr. Zivi, for encouraging me to pursue the Trident program from the first time we met.

I thank my parents, Joe and Rebecca Doherty, for pushing me throughout my academic career. Their love, instruction, and guidance have defined the man and student I have become. Finally, I wish to thank my many friends in company for their assistance throughout this research. Their encouragement never went unnoticed.

Table of Contents

I. Introduction

Problem Statement

It is important for an Unmanned Surface Vehicle (USV) to avoid being identified as a suspicious vessel when conducting Intelligence, Surveillance, and Reconnaissance (ISR) operations in hostile environments. This problem involves the two competing objectives of having the USV travel in a manner that allows it to gather information for its ISR mission while also attempting to blend into the normal maritime traffic to avoid being identified as an intelligence gathering asset. A path planning process can be used to determine effective routes for the USV to follow in order to balance these objectives. In order to collect quality intelligence on the target of interest, the USV needs to follow a path that allows it to get close to the target and remain in the vicinity of the target for an extended period of time. At the same time, the USV must avoid being identified as a vessel exhibiting unusual or suspicious behavior. A path planning method therefore needs to take into consideration how the USV can mimic the behaviors exhibited by the normal traffic. This research develops a path planning process to allow USVs to successfully conduct ISR missions without being identified as intelligence gathering vessels. This research does not address the lower-level control of the USV as it executes its mission or the identification concerns posed by the physical appearance of the USV.

Motivation

Compared to the amount of time spent on the research and development of Unmanned Aerial Vehicles (UAVs) and Unmanned Undersea Vehicles (UUVs), USVs have thus far received less attention. However, USVs provide the Navy with many of the same capabilities as UAVs and UUVs [1]. Unmanned systems in general are effective at completing missions deemed “dull, dirty, or dangerous.” This involves missions that are long and mundane in nature, missions that may expose assets to nuclear, chemical, or biological contamination, and missions that put personnel at a high risk of injury or death [2].

When compared to UAVs and UUVs, USVs have a few important advantages. First, by nature of their operation on the surface, USVs can employ both underwater sensors such as SONAR, depth transducers, and water samplers along with above the surface sensors such as RADAR, infrared sensors, and visible-light cameras. Typically, only the above surface environment or the below surface environment is available to the sensors of UAVs and UUVs respectively. Another significant advantage of USVs over both UAVs and UUVs is their ability to carry a large payload. This larger payload translates to larger sensor suites and more fuel. The third principal advantage of USVs is their capacity for long endurance missions. In comparison to UAVs, USVs are able to loiter in place while expending little energy. UUVs typically must run off of batteries while most USVs burn traditional hydrocarbon fuels. The higher energy density of hydrocarbons gives the USVs greater endurance and speed than the UUVs. USVs also have an advantage over UUVs in the areas of communications and data transfer. Communicating with a vessel on the surface, especially over long distances or at a high data rate, is much easier than with a submerged vessel. The principal disadvantages of using USVs are that they travel at slower speeds than UAVs and are less stealthy than UUVs [1].

One of the most significant challenges to operating a USV, particularly in hostile environments, is the level of autonomy required by the vessel. An autonomous USV differs from a remotely-operated USV in that it does not routinely "reach-back" to its human operator for operational instructions. It is desirable for unmanned systems to operate with a significant amount of autonomy in order to reduce the amount of bandwidth taken up by each unmanned platform. Demands on bandwidth are reduced by minimizing the frequency with which the unmanned system must communicate with its human operator and the complexity of the control signal. As a result, the USV must be designed to conduct its mission and make decisions along its route with limited human involvement [3]. The desire for autonomy is increased when operating in a hostile environment due to the risk a USV incurs when it relies upon two-way transmissions between the USV and its controller. These transmissions increase the likelihood of the USV being detected by enemy forces and provides an opportunity for an enemy to gain control of the USV by executing a "man-in-the-middle" style attack [1].

One control system capable of operating a USV autonomously is NASA's Control Architecture for Robotic Agent Command and Sensing (CARACaS). This control system was originally developed by NASA's Jet Propulsion Laboratory for use with Mars rovers. The CARACaS system is composed of a perception engine, dynamic planning engine, and a behavior engine that allows it to control an unmanned system based on its desired objectives and the nature of its environment. The system is designed to make decisions even when given multiple, competing goals [4]. The Office of Naval Research (ONR) has supported the development of the CARACaS system and successfully demonstrated the system's capabilities in a "swarm style" defense of a Navy vessel in 2014 [5]. More recently, the Navy demonstrated the ability of a swarm of unmanned patrol craft using the CARACaS system to execute a harbor defense mission in late 2016. An attractive feature of the CARACaS program is that a traditionally manned vessel can be converted into a USV with hardware and software additions [6]. This means that operational USVs can be made out of a potentially wide range of existing vessels.

Among the many missions that are suggested for USVs, this research focuses on ISR in hostile environments. Conducting ISR in a hostile environment includes having USVs patrol enemy coastlines or in proximity to enemy surface combatants. The USVs can utilize a variety of onboard and towed sensors to collect intelligence including enemy radio emissions, imagery, and passive SONAR signatures of particular vessels. The USVs can also gather valuable environmental information such as meteorological and bathymetric data to be used in planning future naval operations [1]. CDR Luis Molina, the deputy department head of ONR's Sea Warfare and Weapons Department, reiterated in December 2016 that the Navy envisions using USVs to conduct intelligence, surveillance, and reconnaissance operations [6]. It is important that the USVs are able to avoid being identified during these operations. This need was demonstrated on December 15, 2016 when the Chinese Navy picked up a U.S. Navy UUV. The UUV was fortunately a relatively unsophisticated glider used to collect oceanographic data and the Chinese eventually returned the vehicle [7]. However, this incident stresses the need for a USV to remain unidentified when conducting ISR missions in adversarial waters to successfully accomplish the mission and avoid international incidents.

II. ISR Mission

Objectives and Desired Functionality

The objective of this research is to develop a path planning process to determine which paths a USV should follow when conducting ISR missions in hostile environments. It is assumed that the USV in question is not able to avoid detection by the adversary, so the USV must instead appear to be an innocent member of normal maritime traffic that does not pose a threat to the adversary. The success of this research is determined by the USV's ability to meet the two primary criteria of collecting high quality intelligence on a designated target and avoiding identification by the adversary as a suspicious vessel. The USV's ability to conduct ISR on a target is measured by the USV's distance to the target and the length of time it spends in the target's vicinity. These measurements are compared to measurements obtained from other possible paths to determine how tactically superior the USV's path is compared to other route options. The ability for the USV to blend into the normal flow of maritime traffic is determined by comparing the USV's path to the paths of the normal observed traffic.

This path planning approach is designed to develop a baseline path that could be delivered to an autonomous control system such as CARACaS. The control system is then responsible for executing the mission and deciding the course and behavior of the USV based on the baseline path and other factors such as the sea state and the location of other vessels [4]. The adaptive behavior of the USV during the execution of its mission is not addressed in this research.

Measure of Mission Success

The USV's success at conducting a given ISR mission depends both on the USV's ability to collect information on the target and its ability to avoid being identified by the adversary as a suspicious vessel. The USV's ability to gather information on the target is related to how close the USV can get to the target and how long it can spend in the target's vicinity. The distance from the vessel to the target is among the simplest metrics of the USV's mission effectiveness but also one of the most important. RADAR, SONAR, and electro-optical sensors all experience signal attenuation due to scattering, absorption, and spreading. Each of these sources of attenuation increases with the distance between the sensor and the target [8]. The length of time that the USV is in the target's vicinity is also important as it determines how much imagery the USV is able to collect. The mission success also depends on the likelihood that the USV will be identified during its mission as a suspicious vessel. A high risk of identification means that the adversary is likely to intercept the USV and compromise the mission.

The overall mission success of the USV depends on a combination of the average distance between the USV and the target, the length of time the USV spends in the target's vicinity, and the risk of identification.

The mission success can be improved by decreasing the average distance between the USV and the target, increasing the length of time that the USV spends in the target's vicinity, or decreasing the risk that the USV is identified. However, it is likely that these three metrics of mission success will compete with each other. For example, a path that brings the USV very close to the target or allows the USV to loiter in the target's vicinity will likely incur a higher

risk of the USV appearing suspicious. The best path for the USV to follow therefore depends on the specific desires of the mission commander responsible for the USV. The mission commander may desire an aggressive path that places a premium on the quality of the intelligence and is therefore willing to incur a higher risk of identification. Conversely, the commander may desire a conservative path where minimizing the risk of identification is the highest priority and the quality of the intelligence is not as important. The paths planned through this research are evaluated on the metrics of average distance to the target, time in vicinity of the target, and risk of identification both separately and in various combinations. The mission success of a path depends on the relative importance placed on each of the three mission success metrics based on the particular operation that the USV is tasked to perform.

USV Mission

This research uses various theoretical ISR missions as the test basis to develop and validate an effective path planning approach. Much of the development of the path planning process is completed with the first theoretical ISR mission, but later ISR missions are used to validate the design of the process. The theoretical ISR missions are developed by identifying potential areas of operations and selecting targets on which the USV needs to collect intelligence. Geographic regions are selected that have sufficient records of normal vessel traffic. The targets within these regions are selected based on the complexity of the traffic patterns surrounding the possible targets. A target with a simple surrounding traffic pattern is selected in the earlier stages of the research in order to develop the fundamentals of the path planning process. Targets with more complex surrounding traffic patterns are selected as the path planning process matures. An effort is made to select targets that simulate the location, shape, and size of actual military targets on which one would conduct ISR operations. However, actual high value installations are not used as theoretical targets in order to avoid security concerns. The generic mission assigned to the USV is to collect intelligence on the designated target with a wide range of sensors including RADAR, SONAR, and day/night cameras.

Fort Carroll, located outside Baltimore, MD, is the first theoretical target used in this research. Potential geographic regions and targets are evaluated with electronic, raster nautical charts from the National Oceanic and Atmospheric Administration (NOAA) [9]. Figure 1 shows one of these charts for the area surrounding Fort Carroll. On the chart, land is colored khaki, water shallower than 18 feet is blue, and water deeper than 18 feet is white. Although the raster nautical charts are useful, the coastline data needs to be transferred to MATLAB, the primary software package for this research. MATLAB is a software program designed for engineers and scientists that runs a matrix-based language in order to solve, analyze, and present the solutions to complex computations [10]. The coastline shown on the nautical chart is transferred to MATLAB by downloading shape files of the coastline from the online NOAA Shoreline Data Explorer [11]. The coastline is displayed in a MATLAB figure shown in Figure 2 with Fort Carroll indicated by the magenta marker.

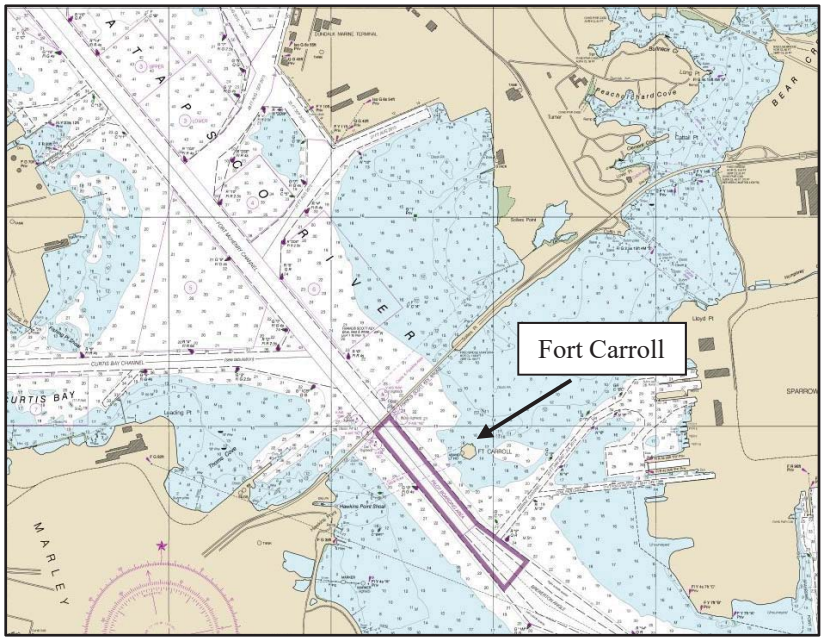

***Figure 1: Raster nautical chart of the area surrounding Fort Carroll* [9]**

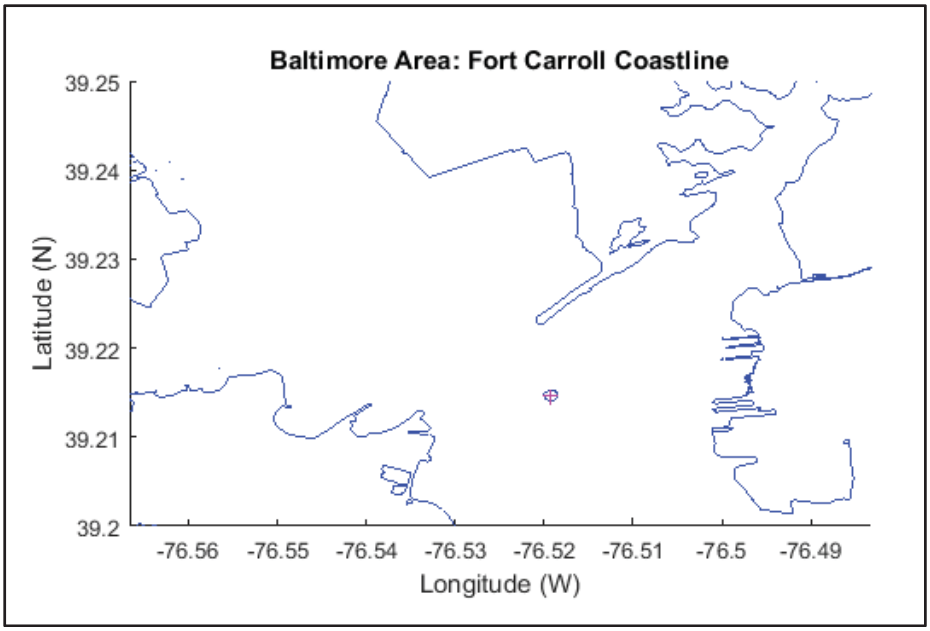

Figure 2: Coastline surrounding Fort Carroll

III. Observed Vessel Traffic

Vessel Traffic Data

Recorded vessel tracks are used in order to determine the normal traffic patterns within a given area. These traffic pattern observations are based on data from the Automatic Identification System (AIS) signals transmitted by many maritime vessels. The International Maritime Organization mandates that most commercial vessels are required to have an AIS system capable of transmitting and receiving vessel position reports. Other vessels can elect to carry AIS systems in order to increase their safety and situational awareness. The AIS system transmits the host vessel's position, course, and speed at an interval of 2-30 seconds depending on the speed of the vessel and the type of AIS transponder. The system also transmits details about the vessel and the vessel's voyage at 6 minute intervals. This information gives a ship's operator advance notice of incoming contacts and a prediction as to how close another ship will pass to his own. The AIS signals are also collected by numerous VHF receivers along the coast maintained by the United States Coast Guard [12].

The historical AIS signals collected by the shore based VHF receivers are publically available through a number of sources. The AIS data used in this research are downloaded from the NOAA maintained website MarineCadastre.gov [13]. AIS data are available from 2009 to 2014 in over 1,200 data sets sorted by geographic region and month. A single data set can contain over 18 million individual AIS signals. The advantage of retrieving AIS data from the NOAA website over other possible sources of data is that the NOAA data are easily accessed and there are ample data sets to work with. This research uses AIS signals from July 2009 in Universal Transverse Mercator Zone 18, the East Coast of the Mid-Atlantic States. July is selected due to the likely high volume of summer boating traffic and 2009 is chosen because it allows specific vessels to be researched based on their MMSI numbers.

Data Sorting

A number of steps are taken in order to filter and sort the raw AIS data into a format that can be easily processed. First, the data from MarineCadastre.gov are downloaded in a geodatabase format that can (to the best of the author's knowledge) only be processed with the ArcGIS software program from software company Esri. ArcGIS is a geographic information system that displays location based data on maps for analysis [14]. The 18 million data points in the July 2009 file are opened in the ArcGIS software as shown in Figure 3. Each AIS signal is shown as a green circle on the map with the darker regions along the coast being a result of the high traffic density in those areas. The geodatabase data is converted to a Comma Separated Value (CSV) file so that the data can be viewed in MATLAB as well as Microsoft Excel and Microsoft Access. It takes approximately 30 hours of processing time to convert the 18 million data points in the geodatabase file to the CSV file format.

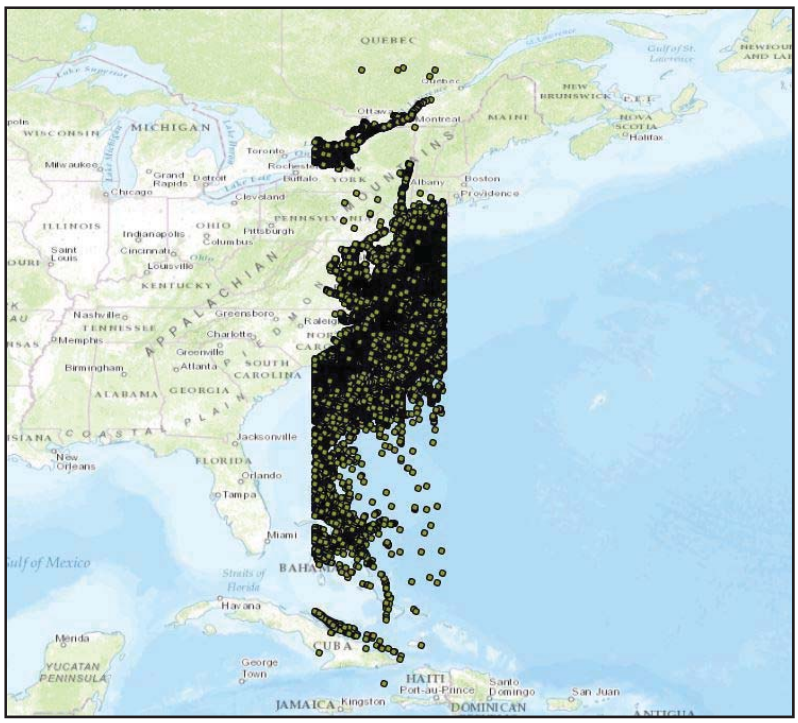

Figure 3: AIS data signals from July of 2009 processed in ArcGIS

Next, a latitude and longitude filter is applied in Microsoft Access to select only those AIS signals located in the area of interest. In the Fort Carroll problem, this area is defined as between 39.183° N and 39.293° N and between 076.413° W and 076.643° W. The data for Zone 18 contains 537,492 data points within the Baltimore area. Each AIS signal contains values in 14 different fields. Many of these fields are not useful for this research, so all AIS signals are streamlined to the following 8 fields: AIS signal ID number, voyage number, MMSI, speed over ground (SOG), course over ground (COG), date and time, latitude, and longitude of the vessel at each AIS transmission. A sample of four streamlined AIS signals is shown in Table 1.

Table 1: Sample of four AIS transmissions from July of 2009

Signal ID	Voyage	MMSI	SOG (knots)	COG	Date and Time	Latitude	Longitude
409268	25	36797970	9	291°	7/1/09 15:34	39.18606°	-076.46059°
409709	361	866891960	6	183°	7/1/09 15:35	39.23700°	-076.53575°
409726	25	36797970	9	291°	7/1/09 15:35	39.18696°	-076.46367°
409729	618	367098250	7	026°	7/1/09 15:35	39.18501°	-076.51440°

The next useful filtering step is to group AIS signals together by vessel. The data downloaded from the NOAA website are organized chronologically by the date and time of the signal's transmission. Rearranging the data by vessel allows one to string consecutive signals from the same vessel into tracks. It is also helpful to remove all of the AIS signals where the vessel's SOG is 0. An AIS system transmits a signal even when the vessel is stationary. Because the signals in

the Fort Carroll problem are for the Baltimore Harbor area, many of the AIS signals are from vessels transmitting while moored alongside a pier. In fact, only 10% of the AIS signals in the Baltimore area are from moving vessels.

The next step is to divide the AIS signals sorted by vessel into separate tracks for each instance of the vessel transiting through the area of interest. This prevents signals from one vessel being strung together into a continuous track if the vessel makes multiple passes through the Baltimore area at different times. Signals from the same vessel are considered to be part of separate tracks if the time difference between consecutive non-stationary signals is greater than 10 minutes. This means that if a vessel enters the Baltimore area and moors alongside a pier, the vessel's departure track is stored separately from its arrival track if the vessel is stationary for greater than 10 minutes. The 10 minute threshold is used because it is not unusual for a few minutes to pass without an AIS signal being recorded for a given vessel due to heavy traffic or radio transmission issues. Figure 4 shows all of the non-stationary AIS signals in the Baltimore area in July of 2009. Figure 5 shows these same AIS signals combined into individual vessel tracks. The traffic density of each of these plots makes it difficult to see individual signals and tracks.

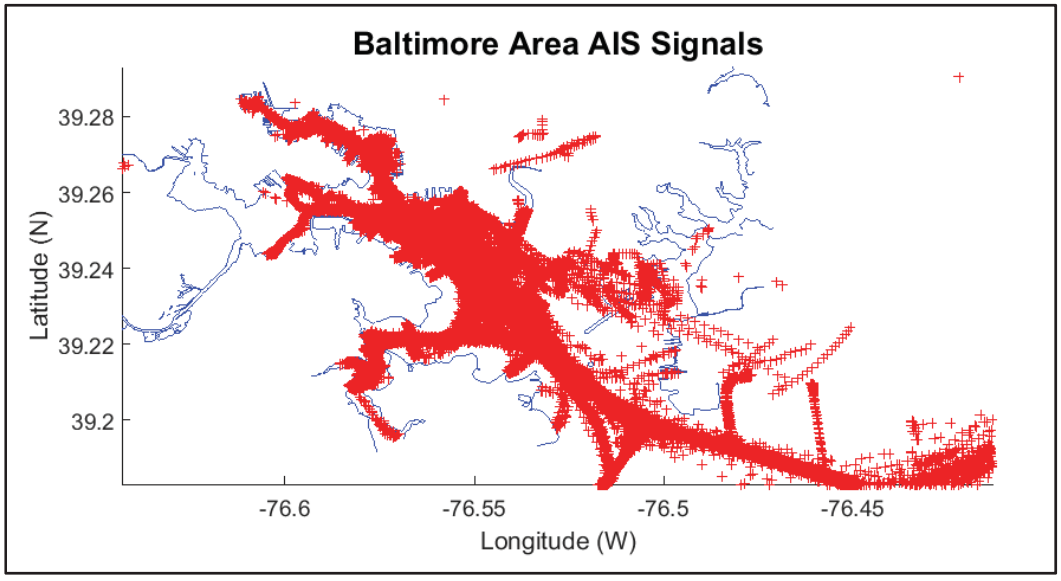

Figure 4: All non-stationary AIS signals in Baltimore area

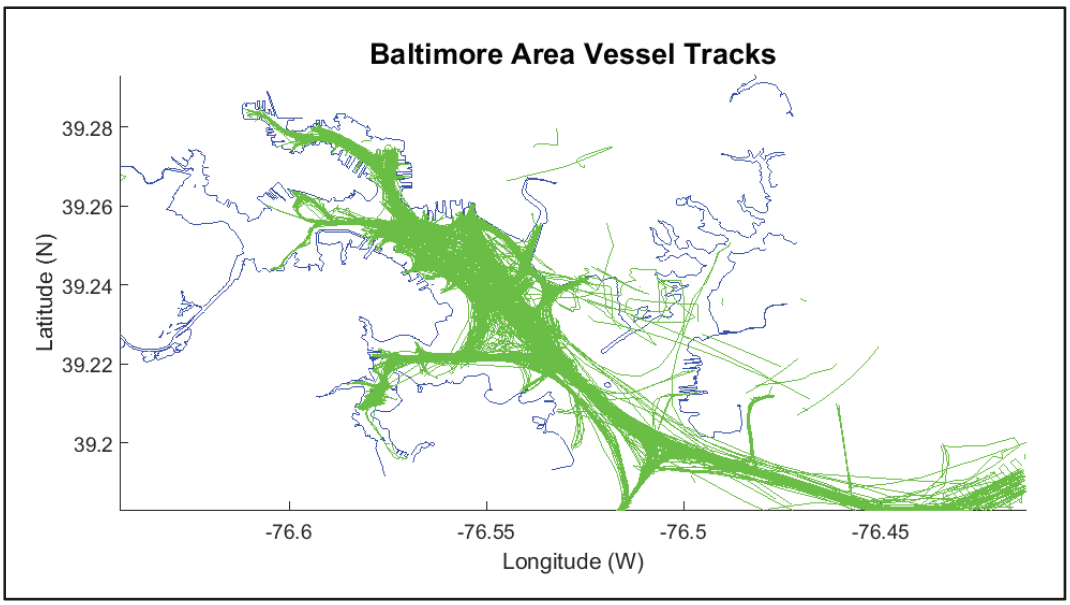

Figure 5: All vessel tracks in Baltimore Area

The sorting described above is completed before any path analysis is performed. However, various portions of the research require additional data filtering and interpolation. Another useful filtering step is grouping the vessel tracks by average heading. This is important because the track that a vessel takes through an area is typically dependent on the direction in which the vessel is heading. This is easily seen in navigational channels where vessels typically stay on the starboard side of the channel similar to cars driving on the right side of a two-way road. It is therefore important that the data used to determine the path for a USV are from vessels travelling in the same direction as that intended for the USV.

Track direction sorting is used to group the vessel tracks passing near the target. The vessel tracks can be sorted based on the COG reported in the AIS signals. However, the AIS signals only provide the instantaneous COG of the vessel at one instance in time. These COG measurements are susceptible to noise from the vessel rocking in waves or the s-turns of an unsteady helmsman. It is more accurate to determine average heading by calculating the azimuth angle between consecutive AIS locations. The average heading of a given track is determined by taking the mean of all of the azimuths between consecutive points within the area of interest. A histogram of the average heading of tracks transiting near Fort Carroll is shown in Figure 6. Based on this histogram, heading bounds can be determined in order to group tracks together based on similar general directions. In this example, there are two main heading groupings from approximately 133° to 161° and from 305° to 334°. The tracks are then sorted into three groups defined as having headings from 123° to 171°, 295° to 344°, and all remaining tracks. Figure 7 shows the tracks in this area differentiated by average course. As discussed above, there is visibly a large difference in track location between the tracks of vessels travelling in a northwesterly direction, shown in green, the tracks of vessels travelling in a southeasterly direction, shown in red, and the remaining tracks shown in magenta.

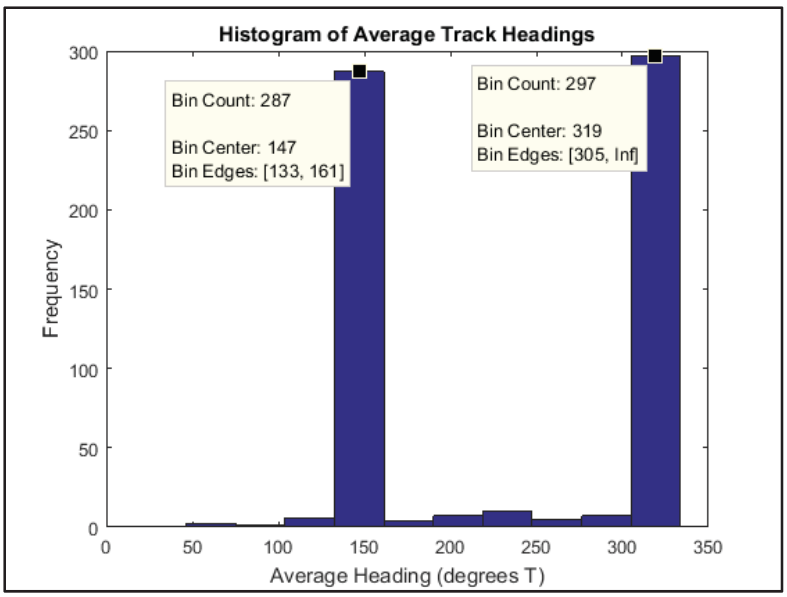

Figure 6: Histogram of average headings of tracks near Fort Carroll

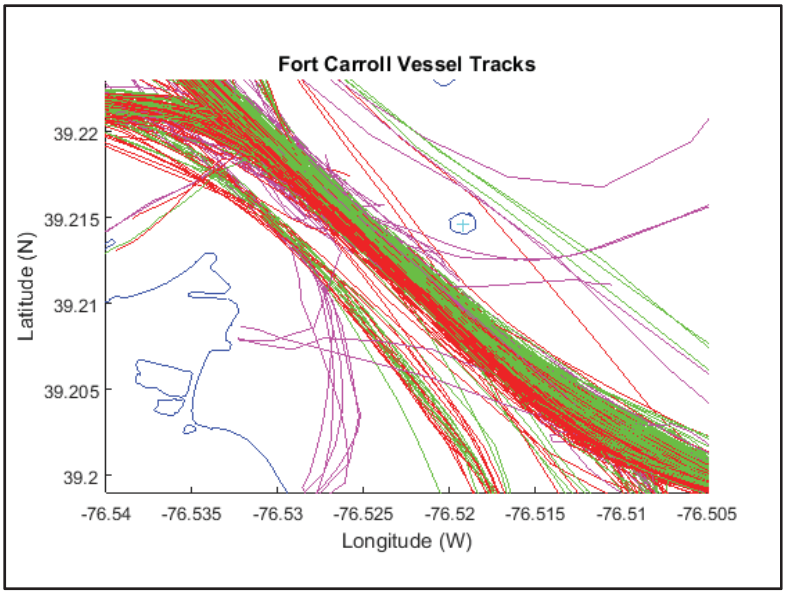

Figure 7: Vessel tracks near Fort Carroll differentiated by average heading

Another data adjustment that is required for some parts of the research is location interpolation between data points. The need for the interpolation is that the AIS signals in the NOAA data are in 1 minute increments. Class A AIS transponders, those found on commercial vessels, transmit signals every 2-10 seconds depending on a vessel's speed and whether it is making significant course changes. Faster vessels transmit with more frequency than slower vessels because faster vessels travel further than slower vessels in the same time frame [11]. A faster vessel therefore requires a faster position update rate in order for the AIS system to maintain a better estimate of the vessel's position between transmissions. However, the NOAA data only has one AIS signal

per minute for each vessel regardless of the speed of the vessel. This is likely to prevent the data sets from becoming larger and more cumbersome to work with. The result of this is that the signal density per unit of path length is higher for slower vessels than it is for faster vessels. An example of this is shown in Figure 8 which displays the tracks of two vessels that travel along similar routes near Fort Carroll. The track on the left is of a vessel travelling with an average SOG of 12.14 knots and the track on the right is of a vessel travelling with an average SOG of 5.93 knots. In the displayed region, the fast vessel has 7 data points and the slow vessel has 114% more data points with 15. This fact is important when conducting analysis on the geographic density of AIS signals. Because there are more data points from the slow vessels than from the fast vessels, the statistical analysis is apt to have a bias towards the location of the slow vessel signals over the location of the fast vessel signals.

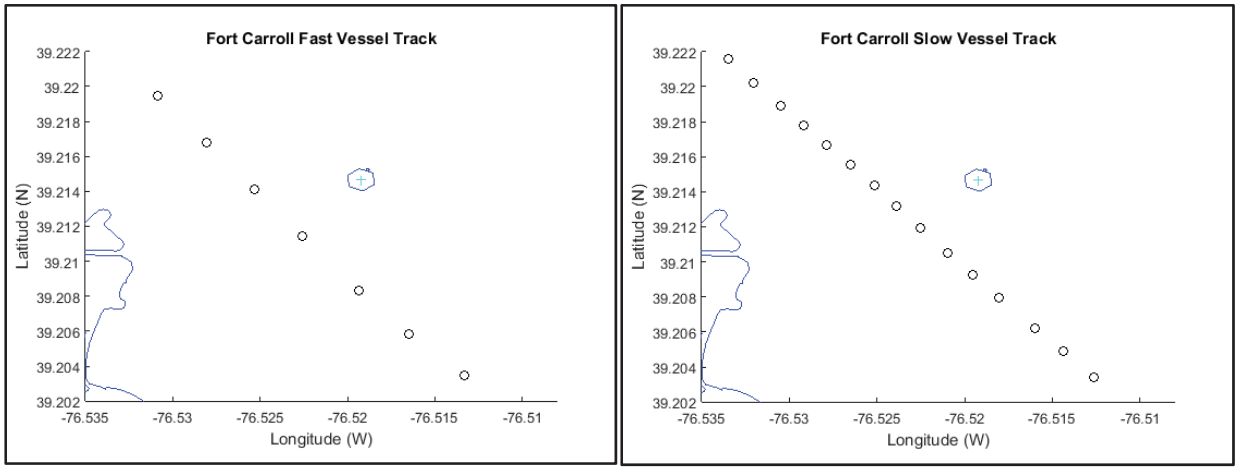

Figure 8: Signal density of vessel tracks at different speeds

The disparity between the signal density of fast vessels and slow vessels is minimized with interpolation. This is done by measuring the distance between consecutive AIS signals for each track. Signals are then added along the linear path between consecutive signals until the distance between all consecutive signals is below a set threshold. For this research, a threshold distance of 100 yards is used. Synthetic AIS signals containing only latitude and longitude information are added to each track so that the distance between one signal and the next is never greater than 100 yards. This interpolation reduces the signal density disparity between fast and slow vessels. The results of the interpolation for the tracks discussed earlier are shown in Figure 9. The interpolated set of data points are marked by the crosses and the original set of data points are marked with the circles. Within the displayed region, the fast vessel track now has 25 data points and the slow vessel track has only 24% more data points with 31. Although this approach does not make all tracks have identical linear signal densities, it provides a significant improvement over using the raw data.

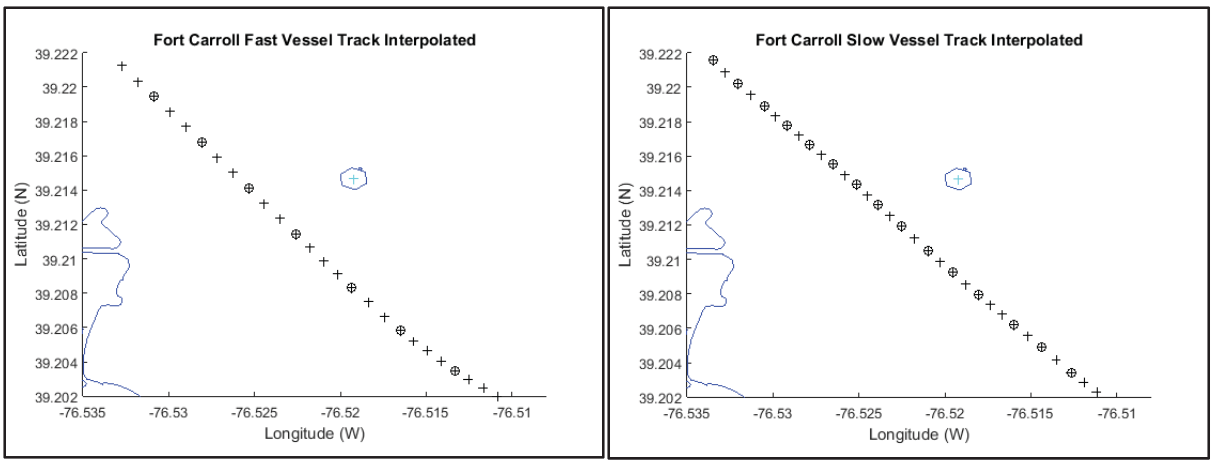

Figure 9: Signal density of interpolated tracks

IV. Vessel Traffic Model

As discussed above, the objective of this research is to develop a path planning approach to allow USVs to blend into normal traffic patterns while conducting ISR operations. This research uses the observed vessel tracks as the foundation of the path planning method. The observed vessel tracks are analyzed in order to determine a plausible path for the USV to follow during its own ISR mission. Three different path selection approaches are pursued during this research. These three approaches are referred to as the "target ray" method, the "surface fit" method, and the "mean path" method. The "mean path" approach is deemed the most suitable method of the three.

Target Rays

The target rays approach selects track segments around the target of interest based on the distance between the track segments and the target. Different track segments are selected over the course of the entire route that continue to allow the USV to maintain a close distance to the target. As shown in Figure 10, one track segment is selected for each sector surrounding the target. These individual track segments form a piece-wise route that serves as the basis for a composite track that the USV can actually follow.

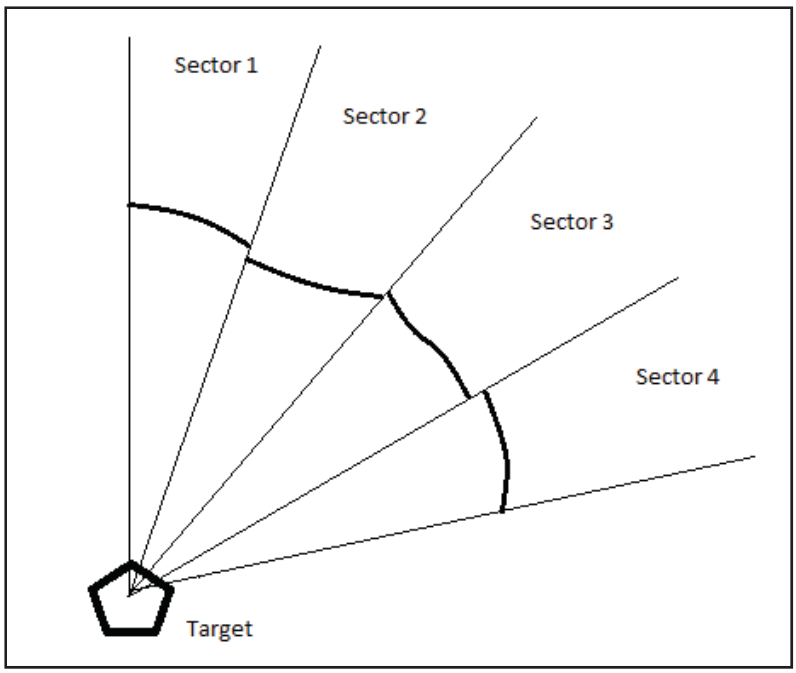

Figure 10: Track segment selection

The distances between the target and the observed vessel tracks are measured along great circle rays radiating from the target. A track's distance from the target is only measured at the intersections between the track and each of the rays. An example of these rays are seen in Figure 11. In this example, rays are generated from 152° to 288° in 4° increments and are shown in black.

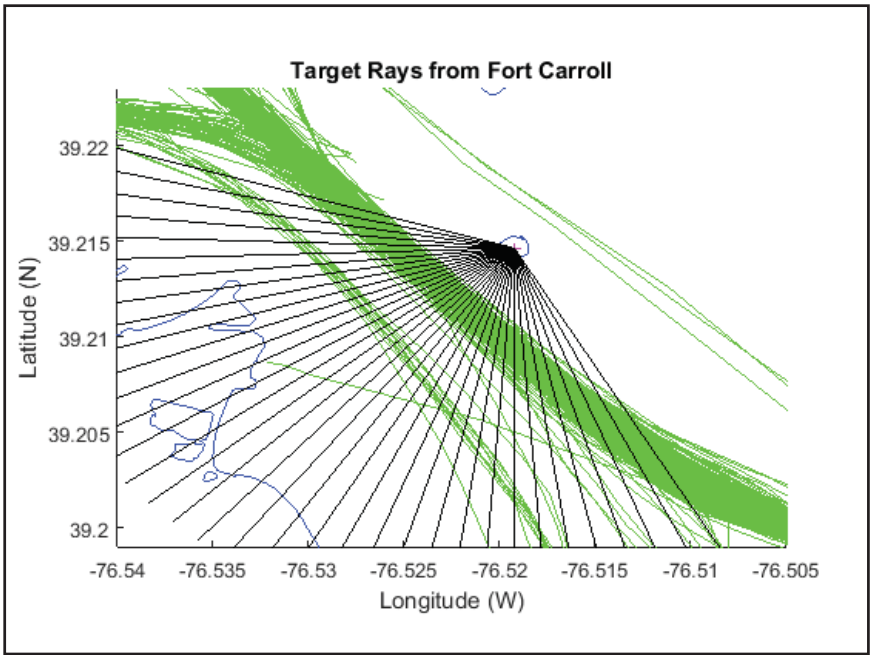

Figure 11: Target rays extending from Fort Carroll

Track segments are selected based on each track's average distance from the target within each given sector. Each sector contains a given number of consecutive rays. An example of a sector is the three rays radiating at 200°, 204°, and 208°. The number of sectors N_s depends on the total number of rays N_r and the number of rays placed in each sector N_{rs} as shown in Eq. (1). The average distance for each track in the sector described above is calculated at the intersection of each of the three rays in this sector with each track. The average of these three distances is used to compare potential track segments for this sector. This relationship is shown in Eq. (2) where $d_{i,j,k}$ is the distance from track k to the target along ray j in sector i and $\overline{D}_{i,k}$ is the average distance between the track and the target along all of the rays in this sector.

$$N_s = \frac{N_r - 1}{N_{rs} - 1} \tag{1}$$

$$\overline{D}_{i,k} = \frac{\sum_{j=1}^{N_{rs}} d_{i,j,k}}{N_{rs}} \tag{2}$$

After the calculation of the average distance between each track and the target within each sector, one can select a track segment appropriate for each sector. This selection is based on the distribution of the average distances between tracks and the target within a given sector, X_A^i. In order to come as close to the target as possible, it is desirable to select track segments that have low average distances. However, selecting the track segments with the lowest average distances may increase the likelihood of the USV appearing as a suspicious vessel. It is therefore important to select track segments that have a low average distance but are also not too close to the fringes of the distribution. A sample distribution of X_A^i for 100 tracks in sector 7, defined by rays 200°, 204°, and 208°, is shown in Figure 12. This histogram indicates that the distribution of average track to target distances is bimodal with peaks between 400 and 500 meters and between 900 and 1,000 meters. This distribution is consistent with what is observed in the tracks displayed in Figure 11 where there is a large concentration of tracks passing close to the target and a smaller concentration of tracks passing approximately twice as far from the target.

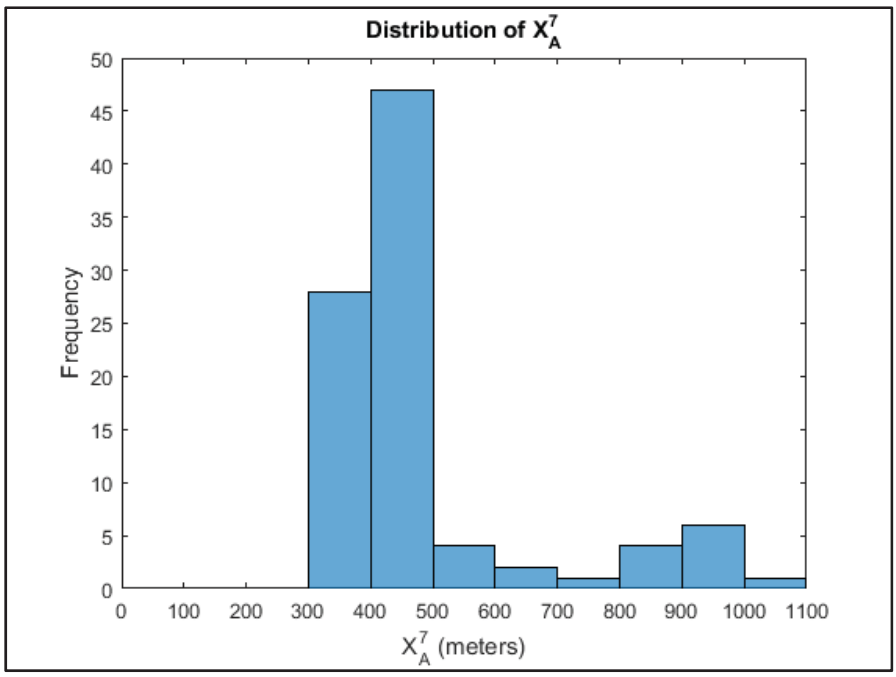

Figure 12: Distribution of distances between target and tracks in sector 7

The distribution of X_A^i for all of the sectors can be viewed in a box plot as shown in Figure 13. This plot indicates the median average distance and the spread between the first and third quartiles, interquartile range, for every sector. Outliers, values that exist more than 1.5 times the interquartile range from the median, are indicated with red asterisks. The large number of outliers is due to the second, further concentration of tracks discussed above.

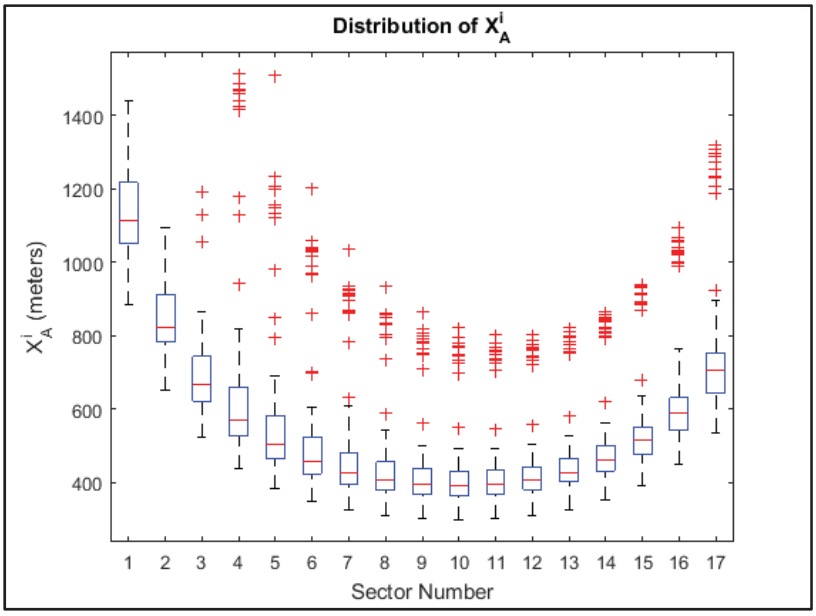

Figure 13: Distributions of target-track distances for 17 sectors

Figure 13 indicates that the interquartile range increases as the average distance between the target and the tracks increases. This is due in part to the angle at which the target rays intersect the observed vessel tracks. As shown in Figure 11, the target rays intersect the tracks closest to the target at approximately 90° angles. However, the angle of the intersection between the target rays and the tracks becomes more acute as the distance at which the rays intersect the tracks increases. The sectors with rays that intersect at the more acute angles generally have larger interquartile ranges because the rays intersect the tracks over longer distances. The relationship between a sector's average track to target distance and the magnitude of the interquartile range heavily depends on the shape of the tracks surrounding the target.

The distributions discussed above are used to help determine the appropriate track segment for each sector. One of the simplest means of selecting track segments for each sector is by selecting all of the track segments that have an average distance at some given percentile from the target. For example, Figure 14 shows all of the track segments at the 10^{th} percentile of the distance from target to track within their given sectors. Some of the consecutive track segments mate smoothly while others do not. These track segments are set to be combined into a continuous, composite path in a later stage.

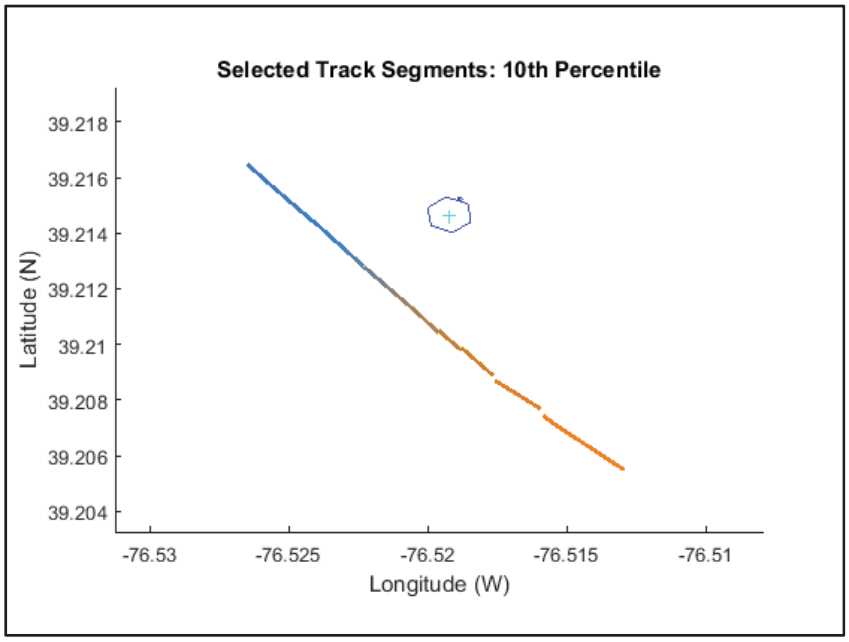

Figure 14: Selected track segments around Fort Carroll

Figure 15 shows the track segments overlaid on the 100 observed vessel tracks. As the track segments in this example are selected at the 10th percentile, they lie closer to the target than 90% of all observed tracks. The three tracks observed above Fort Carroll are not included in the target ray computations.

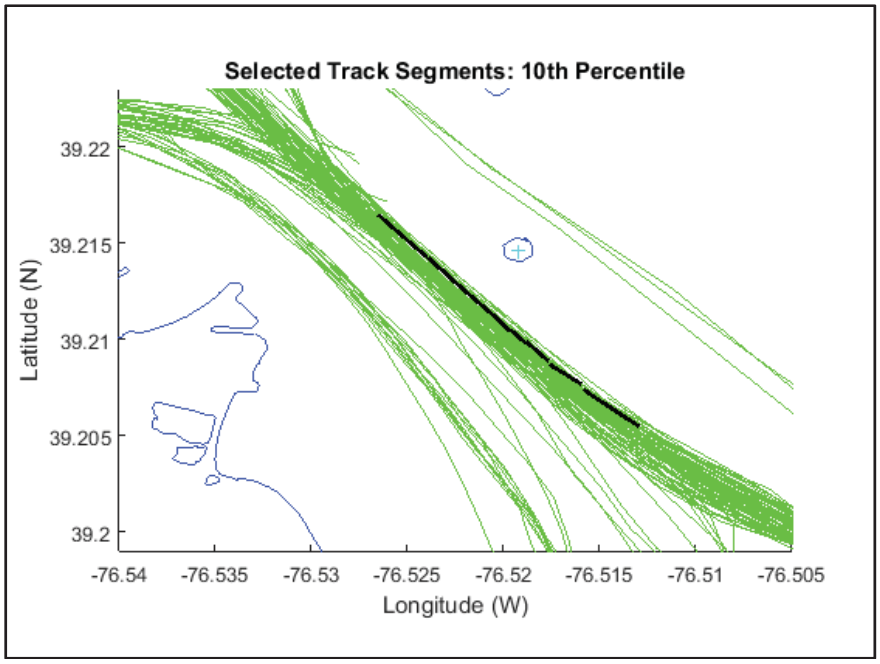

Figure 15: Selected track segments and observed vessel tracks

As shown in the example above, the target rays approach can determine a plausible baseline path for a USV to follow during an ISR mission. The target rays method is well suited for the above example where one is only solving for the USV path in the immediate vicinity of the target and the observed tracks pass by the target in a straight line. However, there are a few limitations with this approach. The most important one is that the target rays approach is not effective at planning paths when the observed vessel tracks travel directly towards or away from the target. In these situations, the target rays intersect the observed tracks at very small angles and can even travel along the observed tracks for an extended distance. Examples of these situations are shown in Figure 16. In this figure, the black pentagon is the target, the green lines represent two observed vessel tracks, the black lines indicate select target rays, and the blue boxes denote areas of concern. In the areas of concern, the target rays approach struggles to determine the average distance between tracks and the target and select appropriate track segments for that sector. Another limitation of the target rays method is that it is not as effective at determining suitable paths at large distances from the target or in complicated traffic patterns where the target rays may intersect tracks multiple times. Figure 16 shows examples of a target ray intersecting a track multiple times in boxes A and B. These situations complicate the track to target measuring process and path segment selection process.

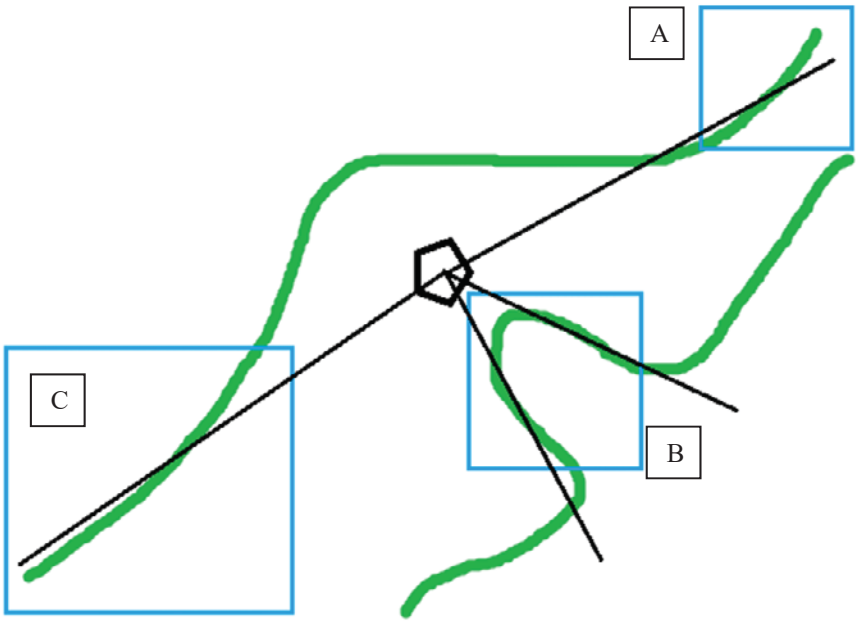

Figure 16: Problem areas with target rays approach

Surface Fit

The surface fit method is an attempt to develop a density map of the AIS signals in the given area of interest. Azariadis performed research in this area by generating a density map of AIS signals in the Mediterranean Sea [15]. The sea was divided into small geographic boxes and the process assigned a value to each box based on the number of signals found inside each box. Paths were then planned through the Mediterranean with a cost function that took into account the value of the boxes that it passed through. This research has an additional step than [15] by attempting to represent the density map with a continuous, differentiable function as opposed to keeping the density function model as discrete weighted regions. A continuous, differentiable function is required in this research due to the optimization methods used to plan paths for the USV.

The first step in the surface fit method is generating a density map of weighted regions similar to that done in [15]. This is accomplished by creating a bivariate histogram of the AIS signals in the region of interest. A bivariate histogram counts the number of times that a data point is within a set range for the first variable and a set range for the second variable. In regards to the AIS signals, the bivariate histogram counts the number of signals that fall within geographic boxes determined by minimum and maximum latitude and longitude limits. Figures 17 and 18 show two orientations of the same bivariate histogram of normal vessel traffic surrounding Fort Carroll. The height of each bin corresponds to the number of AIS signals found within the geographic boundaries of each bin. The channel near Fort Carroll can be identified by the series of tall bins in the center of the image.

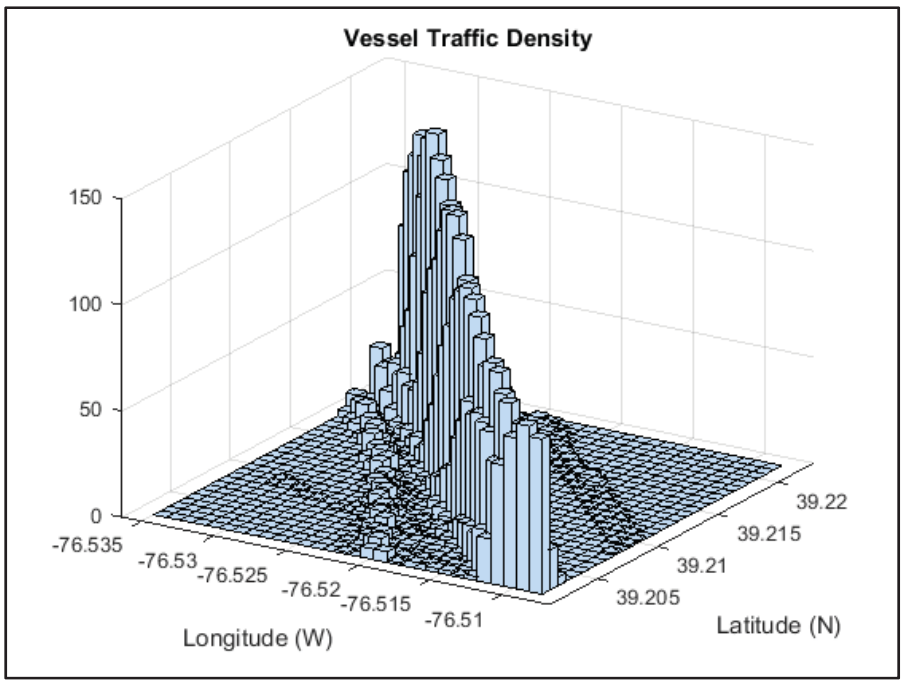

Figure 17: Bivariate histogram view 1

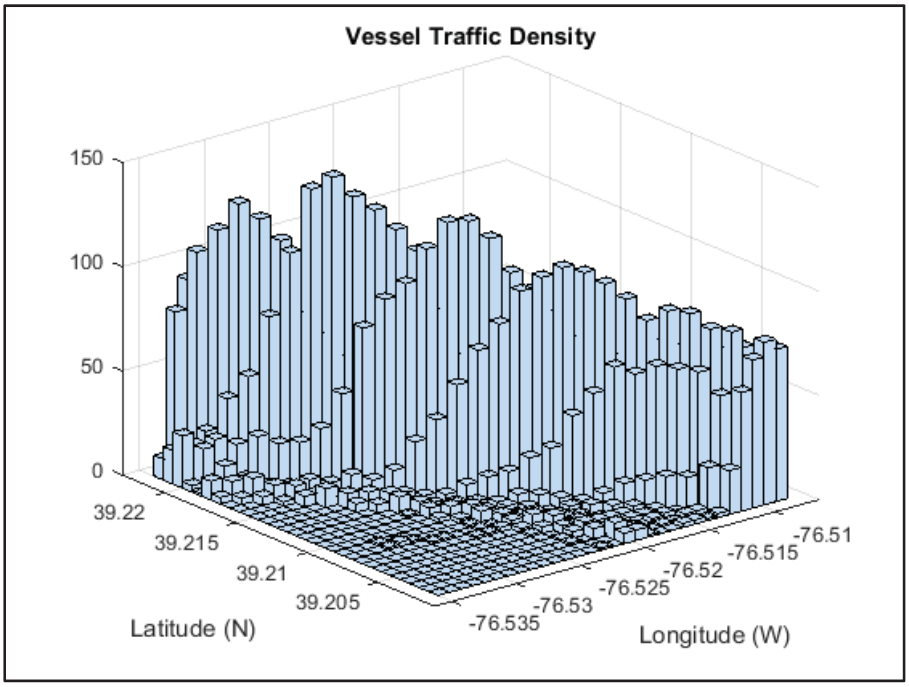

Figure 18: Bivariate histogram view 2

The next step to the surface fit method is generating an equation that can model the density of the AIS signals displayed in the bivariate histogram. An equation is necessary in order to use the density information in later steps of the path planning process. A surface fit equation is designed

to model the number of AIS signals, represented by z, based on a given geographic location, represented by x and y, as shown in Eq. (3). The data for the surface fit takes the form of (x, y, z) triplets where x and y are the center coordinates of a given histogram bin and z is the height of the bin.

$$z = f(x, y) \tag{3}$$

Although other surface fitting methods exist, this research focuses on polynomial regression because of its ability to model a surface with a relatively compact equation and its ease of execution. The regression is completed with the curve fitting toolbox in MATLAB. The regression analysis adjusts the coefficients within a given order polynomial in order to minimize the summed square of the residuals. The MATLAB curve fitting toolbox is limited to a 5th degree polynomial surface fit. Although a surface fit with a higher degree than 5 has the potential to more closely model the data, a higher degree polynomial also can "over-fit" the data or produce very large coefficients. The 5th degree polynomial limitation within MATLAB is therefore not the limiting factor in using this approach as will be discussed later in the report. The general form of the 5th degree polynomial surface is shown in Eq. (4). The regression process solves for the 21 coefficients denoted by p_{xy}.

$$z = p_{00} + p_{10}x + p_{01}y + p_{20}x^2 + p_{11}xy + p_{02}y^2 + p_{30}x^3 + p_{21}x^2y + p_{12}xy^2 + p_{03}y^3 + p_{40}x^4 + p_{31}x^3y + p_{22}x^2y^2 + p_{13}xy^3 + p_{04}y^4 + p_{50}x^5 + p_{41}x^4y + p_{32}x^3y^2 + p_{23}x^2y^3 + p_{14}xy^4 + p_{05}y^5 \tag{4}$$

The initial surface fit for the Fort Carroll example is shown in Figure 19. Visual inspection shows that the developed surface is generally flat and does not adequately model the details of the vessel traffic in the main channel. There are also significant minor ridges and valleys along the surface. Some of the coefficients have values on the order of 10^{15} and the adjusted R-square value for the quality of the fit is 0.2312. This analysis indicates that the initial surface fit is a poor representation of the observed vessel traffic. The surface does not provide a model of the traffic in the main channel with sufficient precision and the minor ridges and valleys introduce unwanted complexity into the model.

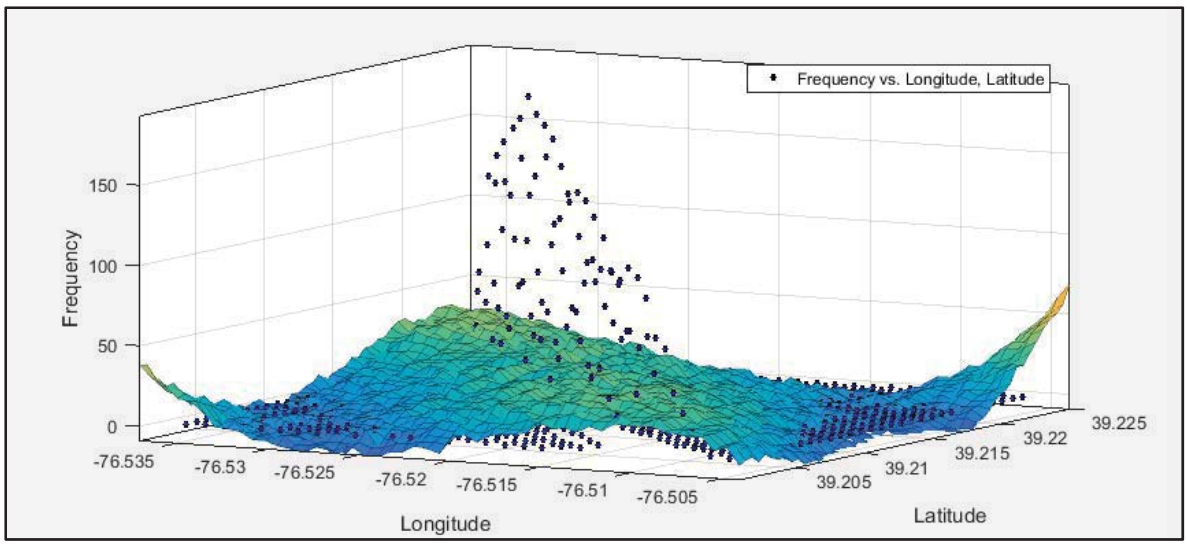

Figure 19: Initial polynomial surface fit

The poor quality of the initial surface fit is improved upon through a few steps. The majority of the geographic area being modeled by the surface does not have any vessels travelling in the northwesterly manner of the filtered data. This results in most of the bins in the bivariate histogram having a bin height of 0. The large number of data points with z values of 0 have a greater influence on the fit of the surface than the smaller number of data points with non-zero z values. This causes the initially modeled surface to be flatter and less precise around the channel than is desirable. Therefore, the data is filtered to remove the data points with z values of 0. As seen in Figures 20 and 21, two orientations of the same fit, this greatly improves the resolution of the surface fit along the center channel. This surface fit is also centered and scaled around (0,0) in the xy-plane to reduce the magnitude of the coefficients. The coefficients reduce in magnitude to the order of 10^2 and the adjusted R-square value improves to 0.3143.

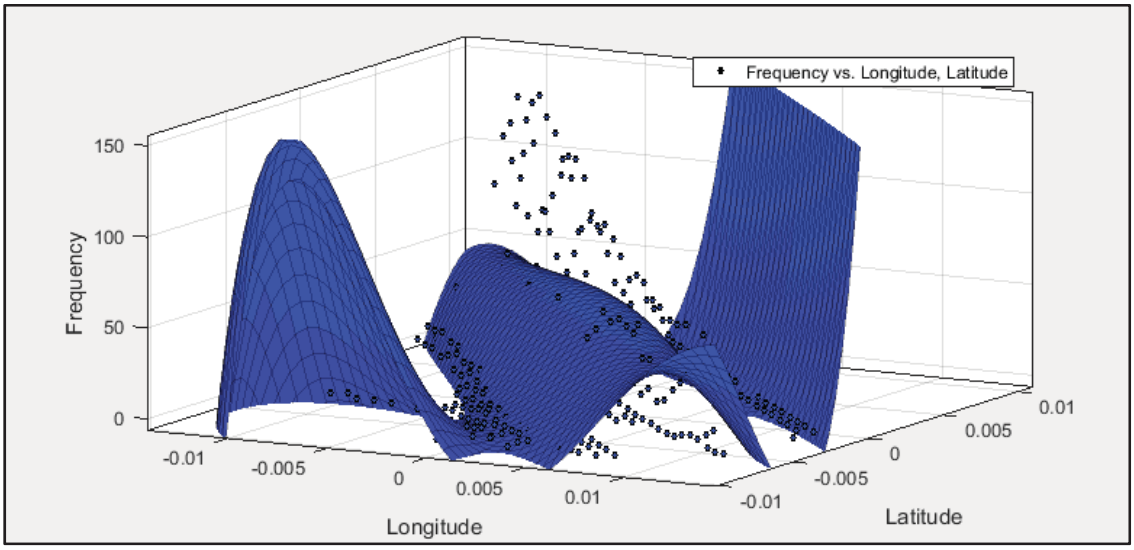

Figure 20: Improved surface fit view 1

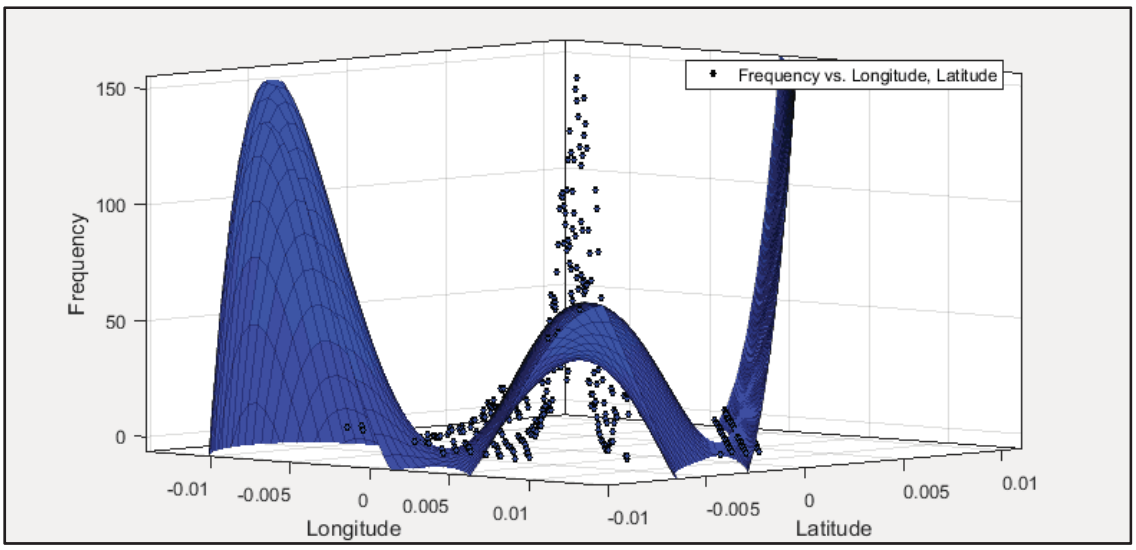

Figure 21: Improved surface fit view 2

Although the surface fit shown in Figures 20 and 21 provides a more pronounced model of the normal vessel traffic along the main channel, the surface model still has a number of problems. The surface is still wider around the center channel than the observed vessel traffic. It is important that the model is able to tightly match the bounds of the main group of vessel traffic because there is a narrow distinction between a vessel being inside the normal traffic lanes and being outside the lanes and potentially appearing as a suspicious vessel. However, the greatest problem with the surface fit is found outside of the main channel. One of the recognized shortcomings of polynomial regression is that it can "over-fit" the data by creating a model that has extreme characteristics as an expense of tightly fitting the data. A side effect of this is that the model can have undesirable behavior outside of the range of the bulk of the data. These issues are clearly discernible in Figures 20 and 21. The few vessel tracks not in the main channel drive the regression procedure to develop a surface that spikes upwards outside of the channel. The extreme behavior of the surface is even more apparent in the zoomed out image of the same surface shown in Figure 22. Although the resulting surface provides a reasonable model of the vessel traffic in the center of the channel, it provides an inaccurate model on the edges of the channel and provides no useful model outside of the channel.

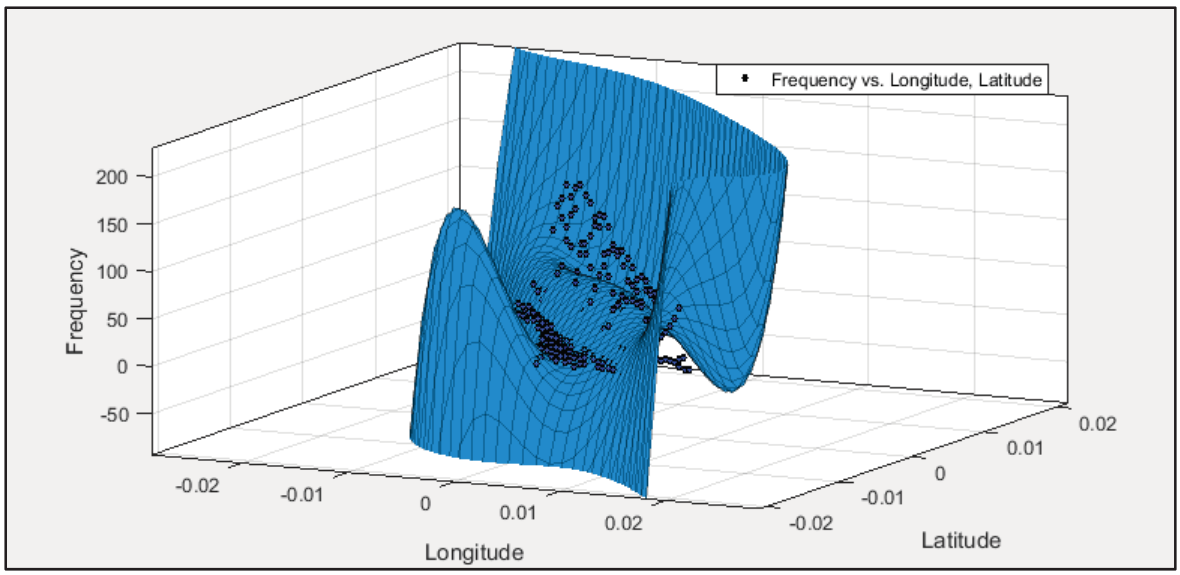

Figure 22: Improved surface fit view 3

Another problem with the surface fit is the magnitude and the number of the coefficients developed by the polynomial regression. Although the magnitude of the coefficients can be reduced by centering and scaling the data as shown in the improved fit, improper scaling of the problem can introduce complications as discussed further below. The mathematical model of the normal vessel traffic is used by the DIDO optimal control software, also discussed in detail further below, in the next stage of the path planning process. DIDO uses numerical methods to arrive at the solutions of a wide range of problems. While these numerical methods are powerful, they can also be sensitive to equations that have large numbers of coefficients and coefficients with large magnitudes. Implementing the above polynomial surface equation in DIDO could result in the software developing an extreme solution with undesirable characteristics.

It is for these reasons that the surface fit method is not considered the best means of mathematically representing the normal vessel traffic. There is the potential to develop a surface fit that accurately models the normal vessel traffic both inside and outside of the channel. However, developing this model is likely to be difficult, require significant data filtering and adjustment, and may still result in a complex surface equation. It is also important to note that the Fort Carroll example is selected because of its simple traffic pattern. The fact that the surface fit method struggles to adequately model the Fort Carroll problem leads one to believe that it is even more difficult to apply the surface fit method to more complicated traffic patterns.

Mean Path

The third approach to modelling the observed vessel traffic is to generate a single path that represents the “average” path through a given geographic region. This mean path is again generated using polynomial regression, but the regression is performed in two-dimensional space vice the three-dimensional regression performed in the surface fit approach. This mean path is then used in the next phase of the path planning process where the “normalcy” of a given path depends on how far the path deviates from the mean path. A path is considered to be less normal the more it deviates from the mean path.

Polynomial regression is normally performed to model the relationship between an independent variable and a dependent variable. The arrangement of the latitudinal and longitudinal coordinates lends itself to declaring longitude as the x variable and latitude as the y variable. Therefore, the (x, y) nature of the AIS signals allows regression analysis to be performed even though neither the latitude variable nor longitude variable can be accurately referred to as the dependent or independent variable. The result of the regression is an equation that estimates a latitude value based on a longitude value. The regression fits a polynomial, a general form of which can be seen in Eq. (5), by adjusting the polynomial's coefficients p_1 to p_{n+1}.

$$y = p_1 x^n + p_2 x^{n-1} + \cdots + p_n x + p_{n+1} \tag{5}$$

The data used in the mean path approach differs from that used in the surface fit approach. The surface fit approach fits a polynomial to the number of AIS signals, z, at various geographic locations (x, y). The mean path approach fits a polynomial directly to the location of the AIS signals.

The polynomial equation is fit to the data using a robust least squares approach. The traditional linear least squares approach fits a polynomial to the data in an attempt to minimize the summed square of residuals. The main problem with this approach is that it makes the mean fit sensitive to outliers. This means that points far from the resulting mean path can have a significant impact on the calculation of the mean path. The robust least squares approach minimizes the impact of these outliers on the resulting mean path. This is accomplished with the bisquare weights method that assigns a weight to each data point [16]. Data points close to the resulting mean path are given a full weight and data points further from the mean path are given progressively smaller weights. Data points far from the mean path are given no weight meaning that they have no impact on the resulting mean path. In the context of the USV problem, this allows the regression to generate a mean path that represents the center channel of the vessel traffic and is not significantly impacted by the various side channels.

The mean path approach is tested on the Fort Carroll example. In this example, a third degree polynomial, of which the general form is shown in Eq. (6), is fitted to the data. Similar to the surface fit approach, selecting the degree of the polynomial that is fit to the data depends on balancing both the quality of the fit and the complexity of the equation. A polynomial of too low of a degree has the potential of not being able to track the details of the observed vessel tracks while a polynomial of too high of a degree generates an equation that is more complicated than is necessary. In the mean path approach, there are no concerns with interpolation outside of the data range because the path is only being planned within the range of the data. The coordinates of the data points are shifted so that they are centered on $(0,0)$ in order to improve the quality of the fit.

$$y = p_1 x^3 + p_2 x^2 + p_3 x + p_4 \tag{6}$$

The result of the third degree robust least squares fit in the Fort Carroll problem is shown in Eq. (7) and Figure 23. This equation is generated with an adjusted R-square value of 0.9884. In the figure, the individual AIS signals, both actual and interpolated, are shown as black dots and the fitted line is shown in blue. Visual inspection indicates that the fitted polynomial is an accurate representation of the vessel traffic along the center channel.

$$y = 811.6x^3 + 5.057x^2 - 0.9049x - 0.0001244 \quad (7)$$

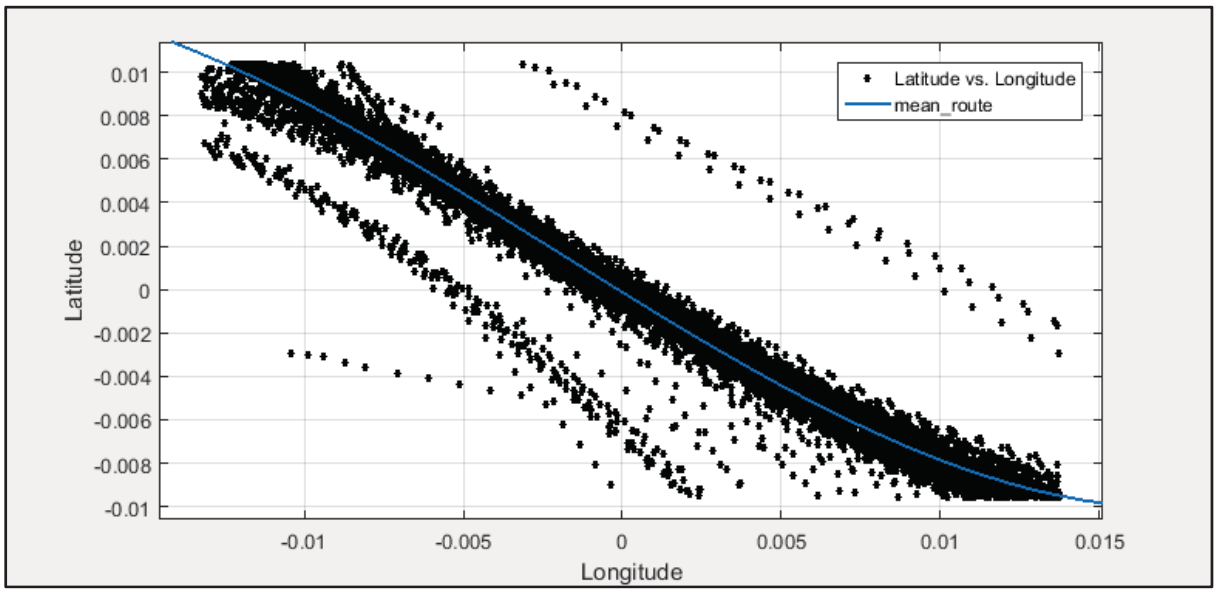

Figure 23: Mean path fit

There are a number of significant advantages to the mean path approach. The first advantage is that the mean path approach is simple to execute. The mean path method only requires minor adjustments to the data after the main filtering and interpolation steps involved in the earlier steps of the path planning process. A polynomial fit for the data is generated in a few minutes by adjusting the desired degree of the polynomial. Another advantage is that the mean path approach mathematically represents the normal vessel traffic in a relatively compact format. The resulting polynomial, although it grows as its degree increases, is simpler than the series of path segments generated in the target rays approach and the surface polynomial from the surface fit approach. The mathematical simplicity of the mean path polynomial is important because it streamlines the operation of the numerical methods in the DIDO optimal control software. The third advantage is that the mean path approach is designed to be resistant to outliers with bisquare robustness. This is important because it allows the mean path approach to focus on representing the bulk of the normal vessel traffic.

The mean path approach also has some limitations. The mean path approach is designed to represent the center of the normal vessel traffic. In most instances, this is likely the center of the starboard side of the channel in whichever direction the traffic is being modeled. The mean path model does not provide any representation of the spread of the normal vessel traffic on either side of the mean path. This is important because the spread of vessel traffic around the center of

the traffic pattern depends on the width of the channel, the presence of shoal water or obstructions, and the number of vessels in the area among other factors. The mean path approach, as currently designed, is unable to represent this spread which is a major shortcoming.

Another disadvantage of the mean path approach is that it has the potential of generating a mean path that no actual vessel would or could follow. The reason behind this is that the polynomial regression is focused on generating an average of all vessel locations along a path and this average path has the potential to fall in water that is not commonly travelled in by normal vessels. Figure 24 shows a theoretical example of vessel traffic dominated by two "railroad track" patterns shown as the thick black lines. The polynomial regression on this traffic pattern is the dashed line that travels in the center of the two "railroad tracks." This mean path is in water that is not normally frequented by any observed vessels.

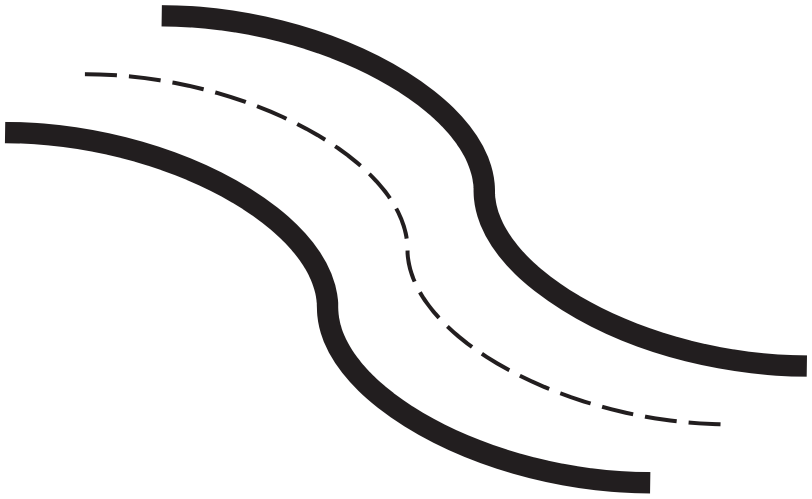

Figure 24: Railroad track traffic pattern

The advantages of the mean path approach in its simplicity in execution and in model representation make it the preferred approach to representing the normal vessel traffic. The limitation of not representing the vessel traffic spread around the mean path is an issue that can be addressed in further research. The risk of generating a mean path that does not exist in nature is a disadvantage that can be avoided with sufficient data filtering.

Speed Selection

The focus so far in modelling the normal vessel traffic has been on representing the path that vessels follow as they transit through the area of interest. It is also valuable to model the speed that the observed vessels travel at through the given area. In order to minimize the risk that the USV appears as a suspicious vessel, it is necessary to match the speed of the USV to the speed of the observed vessels. However, there is also a desire for the USV to maintain a slow speed in order to enhance its ability to gather intelligence on the target. The USV is able to remain in the vicinity of the target for a longer period of time along a set path if it travels at a slower speed.

Speed over ground is one of the data fields in every AIS signal. The average speed of a vessel in a given geographic area can therefore be calculated based on the speed over ground measurements of all of the AIS signals from that vessel in that region. Figure 25 shows a histogram of the average vessel speeds less than 20 knots in the Fort Carroll example. 99% of the observed traffic has an average speed less than 20 knots.

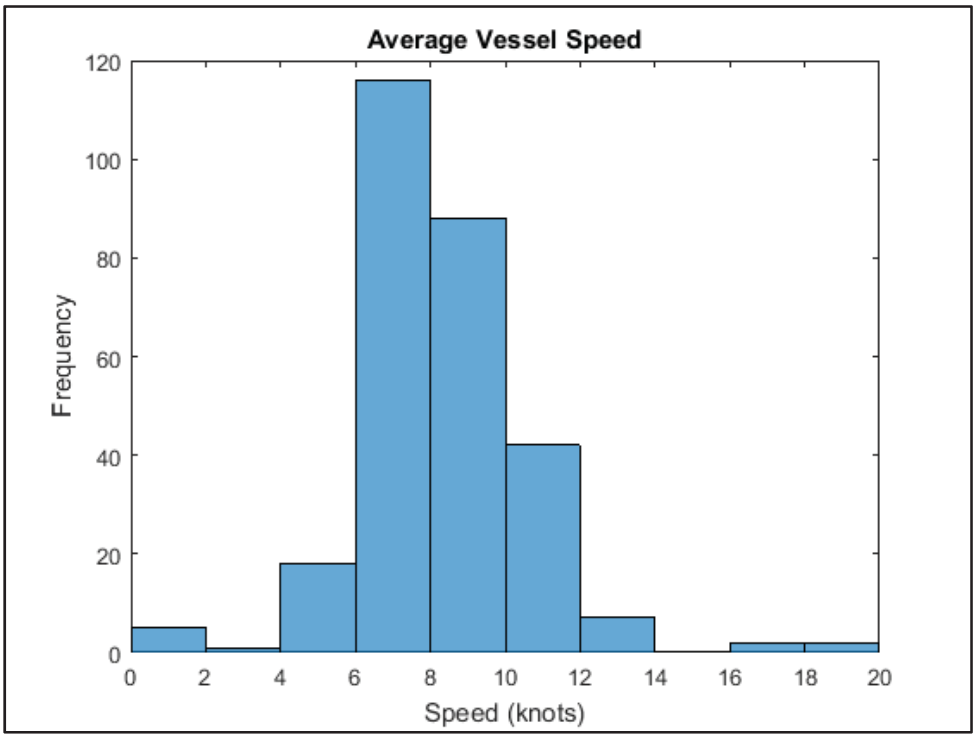

Figure 25: Distribution of vessel speeds

The above distribution of average vessel speeds is used to determine an appropriate speed for the USV. It is possible to vary the speed of the USV along its path, but the USV's speed is held constant in this research. This is done for simplicity but also because most vessels travel at constant speeds for long periods of time. There may be some benefit in varying the speed of the USV but that conclusion is left to further research. An appropriate speed for the USV is selected by calculating the speed that exists at the 10^{th} percentile of the observed vessel speed distribution, 6.0 knots in the Fort Carroll example. This speed is converted to 2.8 $x10^{-5}$ degrees/second to match the units used in the problem framework. This speed is slow so that it provides the USV with more time to collect intelligence, but it is still fast enough to not cause the USV to draw suspicion from the adversary.

V. Selection of an Optimal Path

Mathematical Optimization of a Vessel's Path

The next phase of the path planning process is developing the path that the USV should follow as it conducts its ISR mission. The research up to this point has worked on modelling the behavior of the normal traffic in the vicinity of the target. It is now necessary to identify how the USV can deviate from the normal traffic patterns in order to improve its intelligence collection capability while not appearing suspicious to the adversary. The development of the USV's path is done within the framework of an optimal control problem. This optimal control problem is then solved with DIDO computer software.

Formulating the USV path problem as an optimal control problem requires a number of steps. The purpose of an optimal control problem is to determine the control inputs for a system governed by given dynamics equations in order to bring the system from a beginning state to an end state while minimizing a given cost function [17]. It is first helpful to define the states of the problem $\underline{x}$. These states define the attributes of the USV that are of interest while it is travelling along its path. The states for the USV problem are the vessel's location defined by its latitude and longitude coordinates, the USV's heading which is referenced in a counterclockwise direction from the positive x-axis, and the USV's angular velocity which is the vessel's rate of turn. These states are defined in Eq. (8). The USV is set to travel at a constant speed defined as v. Figure 26 provides a visual of the vessel states x, y, and θ.

$$\underline{x} \in X \subset \mathbb{R}^{N_x}$$

$$\underline{x} = \begin{bmatrix} x \\ y \\ \theta \\ \omega \end{bmatrix} = \begin{bmatrix} longitude \\ latitude \\ heading \\ angular\ velocity \end{bmatrix} \tag{8}$$

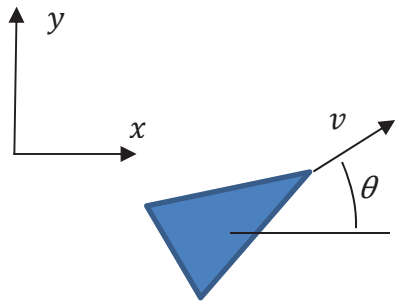

Figure 26: Vessel reference frame

Next, it is necessary to define the control input to the system. The control input, $\underline{u}$, is the decision variable that the optimal control problem is solved for by DIDO. The control input is adjusted so that the states of the problem minimize the given cost function. The control input for the USV

problem is the vessel's angular acceleration α as shown in Eq. (9). Angular acceleration is used as the control for the USV in order to generate a smooth path even if the control solution has rapidly changing values.

$$\underline{u} \in U \subset \mathbb{R}^{N_u}$$

$$\underline{u} = [\alpha] = [angular\ acceleration] \tag{9}$$

In order to reduce the time it takes for the software to arrive at a solution and to prevent unrealistic or undesirable solutions, it is necessary to place bounds $\underline{h}$ on the solution space. This means placing lower, $\underline{h}^L$, and upper bounds, $\underline{h}^U$, on the state and control variables. In relation to the vessel's location, this means keeping the USV within a set geographic box. The vessel's heading is also constrained to prevent the USV from making loops along its path. The bounds on the angular velocity and angular acceleration are set to realistic values for a vessel. The state and control variable bounds are shown in Eq. (10).

$$\text{General form: } \underline{h}^L \leq \underline{h}(\underline{x}(t), \underline{u}(t), t) \leq \underline{h}^U \tag{10}$$

$$\begin{gathered} x_{min} \leq x \leq x_{max} \\ y_{min} \leq y \leq y_{max} \\ \theta_{min} \leq \theta \leq \theta_{max} \\ \omega_{min} \leq \omega \leq \omega_{max} \\ \alpha_{min} \leq \alpha \leq \alpha_{max} \end{gathered}$$

With the state and control variables defined, it is now possible to define the dynamics of the USV system, Eq. (11). The dynamics define how the USV travels from beginning state to end state. The equations of motion for the USV problem are simple two-dimensional kinematics defining the first derivative of the state vector, $\dot{\underline{x}}$, over time, t, as being equal to relationships with the speed constant, state variables, and the angular acceleration control input.

$$\dot{\underline{x}}(t) = \underline{f}(\underline{x}(t), \underline{u}(t), t) = \begin{bmatrix} \dot{x} \\ \dot{y} \\ \dot{\theta} \\ \dot{\omega} \end{bmatrix} = \begin{bmatrix} v\cos\theta \\ v\sin\theta \\ \omega \\ \alpha \end{bmatrix} \tag{11}$$

Another part of the optimal control problem is the endpoint constraints, $\underline{e}$. These constraints define the starting condition, $\underline{e}^L$, and ending condition, $\underline{e}^U$, of the vessel's state and control variables. The endpoint constraints define specific values or a range of values for each of the initial and final state variables and the time variable. In the USV problem, the most important endpoint constraints are the starting location and ending location of the USV path. These beginning and ending locations are set to require a solution path that brings the USV through the waterway near the target of interest. Looser constraints are placed on the other state and control variables to help bound the solution space. The endpoint constraints are shown in Eq. (12) with t_0 and t_f denoting the starting time and ending time respectively.

$$\text{General form: } \underline{e}^L \leq \underline{e}(\underline{x}_0, \underline{x}_f, t_0, t_f) \leq \underline{e}^U \tag{12}$$
$$x(t_0) = x_{start}$$
$$y(t_0) = y_{start}$$
$$\theta_{L,start} \leq \theta(t_0) \leq \theta_{H,start}$$
$$\omega_{L,start} \leq \omega(t_0) \leq \omega_{H,start}$$
$$x(t_f) = x_{end}$$
$$y(t_f) = y_{end}$$
$$\theta_{L,end} \leq \theta(t_f) \leq \theta_{H,end}$$
$$\omega_{L,end} \leq \omega(t_f) \leq \omega_{H,end}$$

The next portion of the optimal control problem is the cost function. There are potentially an infinite number of paths from the beginning point to the end point which follow the vessel dynamics and are within the state and control variable bounds. The cost function determines which of these paths best satisfies its conditions. The cost function, J, is an expression of what characteristics of the path or behavior of the vessel are most desirable [17]. It can be calculated from a function of the endpoint states, expressed as E, and a function of the running states, F, as shown in Eq. (13).

$$J[\underline{x}(\cdot), \underline{u}(\cdot), t_0, t_f] = E(\underline{x}_0, \underline{x}_f, t_0, t_f) + \int_{t_0}^{t_f} F(x(t), u(t), t)\, dt \tag{13}$$

For the USV problem, a path is desired that allows the USV to collect as much intelligence as possible on the target while allowing the USV to blend in with the normal traffic patterns. The solution to the USV problem is the path that minimizes the cost function while adhering to all of the constraints. The USV cost function only involves the running cost component. The quality of the intelligence that the USV is able to capture depends on the length of the time that the USV remains in the vicinity of the target and how close the USV is able to come to the target. The length of time that the USV remains in the vicinity of the target can be increased by either increasing the length of the path in the vicinity of the target or decreasing the USV's speed. It is difficult to dramatically increase the length of the USV's path without having the USV execute S-turns or multiple passes in the channel, both of which are activities likely to draw suspicion from the adversary. Therefore, decreasing the USV's speed is the better option. A low speed for the USV is selected from the distribution of observed vessels as discussed above. As the USV holds a constant speed over the duration of its path, the focus on improving the intelligence gathering capability of the USV is on decreasing the distance between the USV and the target. The cost function therefore computes the distance between the USV and the target as the USV travels along its path. The cost function integrates these distances over the duration of the USV's mission. A lower total cost corresponds to the USV drawing closer to the target throughout its path and therefore having a better intelligence gathering capability. This is expressed in Eq. (14). The explanation of how this distance calculation is made is discussed further below.

The cost function also needs to measure how well the USV is able to blend in with the normal traffic patterns. The mean path approach provides a single path that theoretically represents the behavior of an average vessel as it passes through the given geographic area. It is therefore possible to represent how well the USV blends in with the normal traffic by measuring the distance that the USV's path deviates from the mean path. This distance is integrated over time

for the duration of the USV's mission. A lower total cost corresponds to a USV path that sticks closer to the mean path and a higher cost corresponds to a USV path that deviates more from the mean path. This relationship is expressed in Eq. (14). A more detailed explanation of how these measurements are made is provided below.

$$Minimize\text{: } J = \int_{t_0}^{t_f} (USV\ to\ target\ distances)\, dt + \int_{t_0}^{t_f} (USV\ to\ mean\ path\ distances)\, dt \tag{14}$$

DIDO Software

It is beneficial at this point to discuss the software that is used to solve the USV problem. This is important because much of the framework of the problem, particularly the cost function, is tailored to the software. The software used in this research is a third-party MATLAB application titled DIDO from software company Elissar Global. DIDO uses pseudospectral optimal control algorithms to solve a wide range of optimal control problems. DIDO allows users to input a problem's cost function, Eq. (13), a body's dynamics, Eq. (11), and a problem's constraints, Eq. (10) and Eq. (12) [18]. The software then calculates an optimal solution based on these factors and the user's desired level of accuracy. The pseudospectral method that DIDO uses divides the optimization problem into a series of smaller optimization problems at a user specified number of nodes. DIDO then approximates the control, state, and co-state variables using Lagrange interpolating polynomials [19]. DIDO refines its approximations down to the final solution by iterating the optimization problem thousands of times. In the context of the USV problem, each node of the solution refers to a coordinate pair along the USV's path. Having the software solve the problem for fewer nodes decreases the solution's accuracy, but it also decreases the amount of time that it takes for the computer to process the solution. A highly accurate solution is produced by inputting a lower node solution as the guess to the answer of a higher node solution of the same problem. The guess decreases the software's required processing time. This process is iterated by progressively increasing the number of nodes and using each solution as the guess for the next iteration until the desired accuracy is achieved.

DIDO Execution of Path Problem

After the paper formulation of the optimal control problem, the next step is implementing the problem in the DIDO software. DIDO is easy to work with in terms of entering the optimal control problem framework into the software. The state and control variables, the bounds for the variables, the vessel dynamics, and the endpoint constraints are entered into separate MATLAB files to be referenced by the software. The greatest challenge with executing the USV path problem is in the implementation of the cost function.

As discussed above, DIDO solves the path planning problem at a set number of nodes between the given starting states and the finishing states. In terms of the USV problem, the solution generated by DIDO is therefore not a solid path but rather a set of consecutive points that can be strung together to form a path. The cost function is integrated over the duration of the path to form a total cost. DIDO iterates the process by adjusting the nodes and reevaluating the total cost. The process iterates until the total cost is minimized. Figure 27 shows a model of a USV problem with the objective of having the vehicle travel from $(0,0)$ to $(12,12)$. An example mean path is generated with the equation $y = x^2/12$ and shown in orange. The black circle indicates the target's location and the blue asterisks represent the DIDO solution nodes.

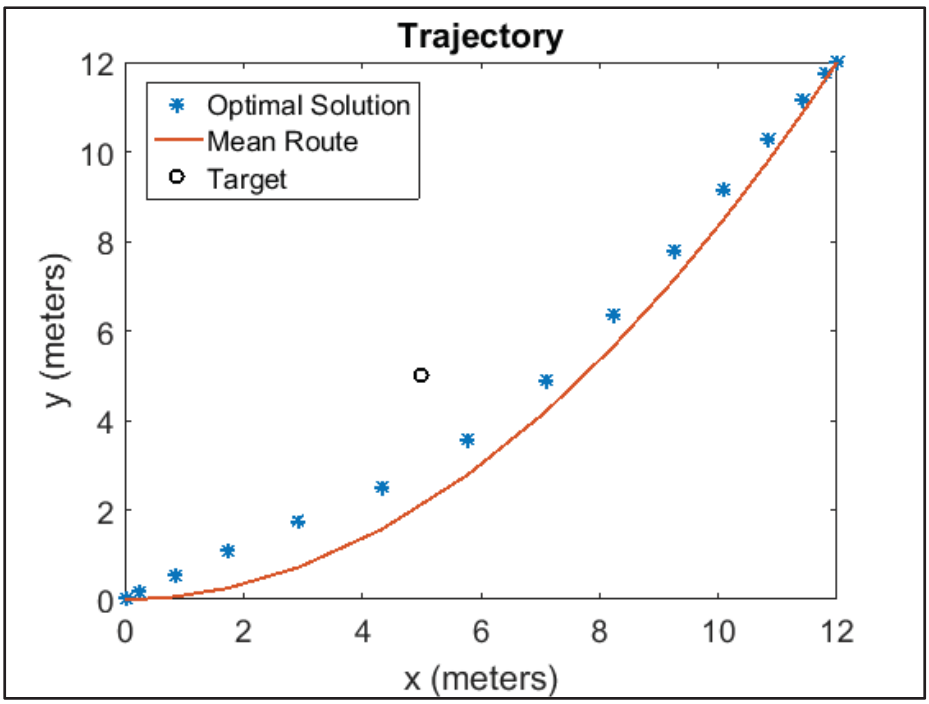

Figure 27: USV problem

A general form of the cost function is shown in Eq. (14). It is difficult to execute the cost function, particularly in regards to the distance between the USV and the mean path, in a manner that DIDO can process efficiently. The distance between the USV's path and the target is a simpler measurement to represent because the target is represented by a single, stationary point. The portion of the cost function responsible for measuring the distance between the USV and the target calculates the Euclidean distance between the target and each of the nodes along the USV's path and integrate these distances over the duration of the mission. The Euclidean distance equation shown in Eq. (15) shows how the distance between a single node and the target can be calculated where $(x(t), y(t))$ represent the coordinates of the node and (x_T, y_T) are the coordinates of the target. This calculation provides the measurement between a single blue asterisk and the black circle shown in Figure 27.

$$Distance = \sqrt{(x(t) - x_T)^2 + \left(y(t) - y_T\right)^2} \tag{15}$$

The individual distances between the nodes and the target are then integrated in order to measure the overall cost between the USV's path and the target. The actual distance between each node and the target does not carry great significance. It is therefore more practical to integrate the square of the distances between the nodes and the target to improve computational efficiency. Eq. (16) performs this computation by calculating the distance between each node in the interval from t_0 and t_f to the target. This computation represents the cost associated with the quality of the USV's intelligence with a lower cost being preferred.

$$J_{USV\ to\ target} = \int_{t_0}^{t_f} [(x(t) - x_T)^2 + (y(t) - y_T)^2] dt \tag{16}$$

The second portion of the cost function is associated with evaluating the distance between the USV's path and the mean path. As mentioned above, this calculation is more complicated because the USV's path is not being compared to a stationary point as is the case with the distance between the USV's path and the target. There are a couple of ways to calculate the distance between each of the solution nodes and the mean path equation. The distance between a solution node and the mean path can be calculated with Eq. (17) where $g(x)$ represents the equation for the mean path and $(x(t), y(t))$ represent the coordinates of a given node.

$$Distance = \sqrt{(x(t) - x)^2 + \big(y(t) - g(x)\big)^2} \tag{17}$$

The shortest distance between the node and the path can be determined by taking the first derivative of Eq. (17) with respect to x and setting it equal to 0. The roots of this function can then be evaluated to determine the point along the mean path that is closest to the solution node. The distance between the node and the closest point on the path can then be calculated with the Euclidean distance equation. The advantage of this approach is that it provides the exact distance between each node and the mean path. The major disadvantage of this approach is that it is computationally demanding. This approach requires the software to compute derivatives, evaluate potential roots, and calculate distances between every node and the mean path for every DIDO iteration. These calculations must be performed for tens of thousands of iterations on potentially high order polynomials. This causes the DIDO software to run inefficiently and take a long time to generate a solution.

The alternative calculation used in this research measures the vertical distance between each solution node and the mean path. This computation is performed by solving the mean path equation with the x-coordinate of a given solution node. The distance between the solution node and the mean path equation is represented as the difference between the y-coordinate of the solution node and the solution of the mean path equation. This difference is squared in order to ensure a positive value. Equation (18) shows how this computation is performed in order to determine the cost of the distance between the USV's path and the mean path over the interval from t_0 to t_f.

$$J_{USV\ to\ path} = \int_{t_0}^{t_f} \Big(y(t) - g\big(x(t)\big)\Big)^2 dt \tag{18}$$

The complete cost function is a weighted combination of Eq. (16) and Eq. (18) and is shown in Eq. (19). The W in the cost function is a weighting variable to determine the relative importance placed on the cost of the distance between the USV's path and the mean path and the distance between the USV's path and the target. This weighting variable is adjusted to determine the desired quality of intelligence and risk of identification. The adjustment of W plays an integral role in the path planning process. The results of adjusting the weighting variable are discussed further below. The variable y_m is introduced to replace $g\big(x(t)\big)$ as the variable representing the y-coordinate of the mean path.

$$J = \int_{t_0}^{t_f} [(y(t) - y_m)^2 + W((x(t) - x_T)^2 + (y(t) - y_T)^2)]dt \tag{19}$$

The principal advantage of representing only the vertical distance between the USV's path and the mean path is computational efficiency. This computation does not require the software to compute derivatives or evaluate multiple roots. One disadvantage of this approach is that it does not calculate the true distance between the USV's path and the mean path. However, the focus is on having a metric to evaluate the distance between the two paths so not calculating the true distance is not a significant shortcoming.

Another necessary step in implementing the optimal control problem framework into DIDO is ensuring that the problem is properly scaled. DIDO relies on numerical methods to calculate optimal solutions. These numerical methods can be sensitive to the scale of the values used in the problem. The software operates most efficiently if the values of the states, control variables, and cost are not exceptionally small, exceptionally large, or vary greatly in order of magnitude.

Fort Carroll Solution

Once the generic optimal control problem is established, the next step is applying this framework to the Fort Carroll example problem. Initial trials with the Fort Carroll problem indicate that the problem requires scaling in order to execute properly due to the reasons discussed above. The scaling is performed by multiplying the centered latitudinal and longitudinal coordinates of the observed vessel traffic by a factor of 1,000. This in turn scales the mean path equation, shown in Eq. (20), and the vessel's speed to 0.027 degrees/second.

$$y = 0.0008116x^3 + 0.005057x^2 - 0.9049x - 0.1244 \tag{20}$$

The Fort Carroll problem is then entered into DIDO with the scaled mean path equation and scaled variable bounds. Figure 28 shows a 30 node solution to one form of the Fort Carroll ISR problem. In this problem arrangement, the weighting factor, W, for the cost function, Eq. (19), is set to 0.25. The solution nodes indicate that the USV's path stays close to the mean path with a deviation in the direction of Fort Carroll. This result visually matches expectations of a potential USV path.

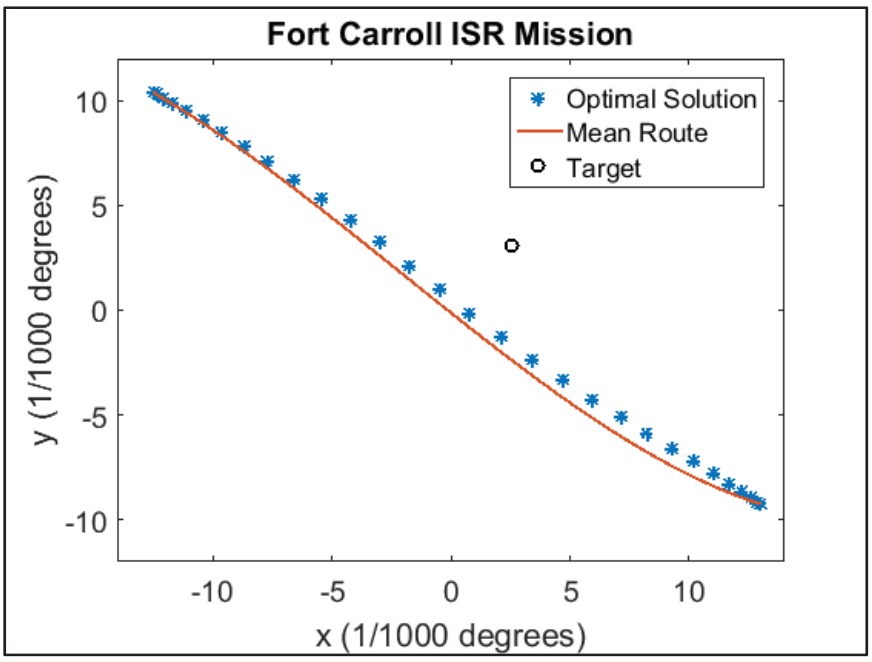

Figure 28: Fort Carroll mission

It is also beneficial to view the generated USV path overlaid on the observed vessel tracks as seen in Figure 29. The thick black line is the generated USV path and the green lines are the observed vessel tracks. Based on visual inspection, the USV's path remains within the main body of the observed vessel tracks throughout its duration. The USV's path also passes closer to Fort Carroll than many of the observed paths.

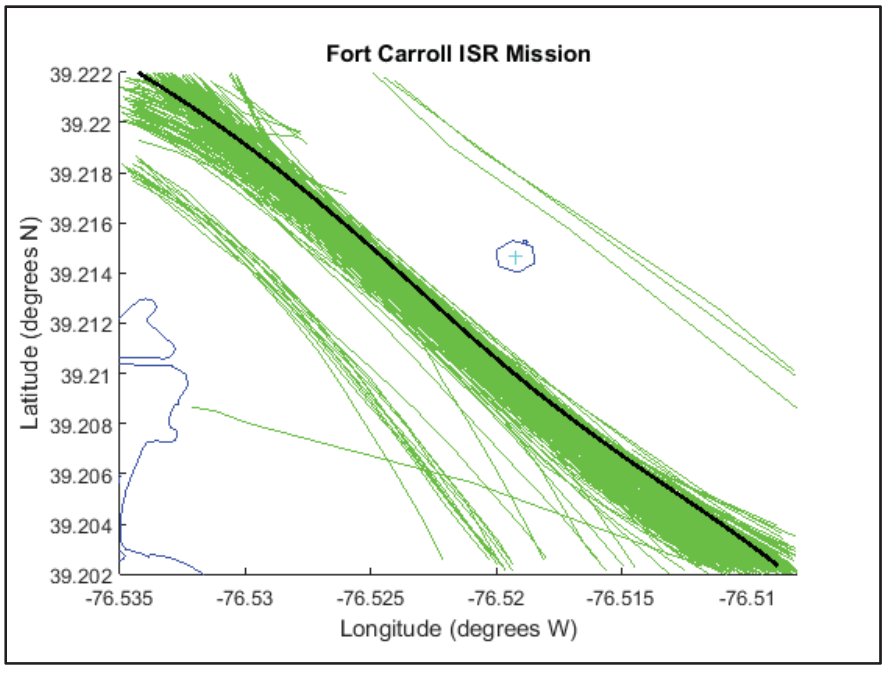

Figure 29: Fort Carroll mission with observed vessel tracks

The aggressiveness of the USV's path, in regards to the level of risk accepted in order to gain higher quality intelligence, can be adjusted by varying the weighting factor. Figure 30 shows the original path, shown in black, developed with a weighting factor of 0.25 along with two additional USV paths developed with weighting factors of 0.5, shown in red, and 1.0, shown in blue. The aggressiveness of the USV's path is expected to increase as the weighting factor increases. The results in Figure 30 match this expectation with the red path drawing closer to Fort Carroll and the blue path drawing closer still. The full path evaluation process will be discussed further below.

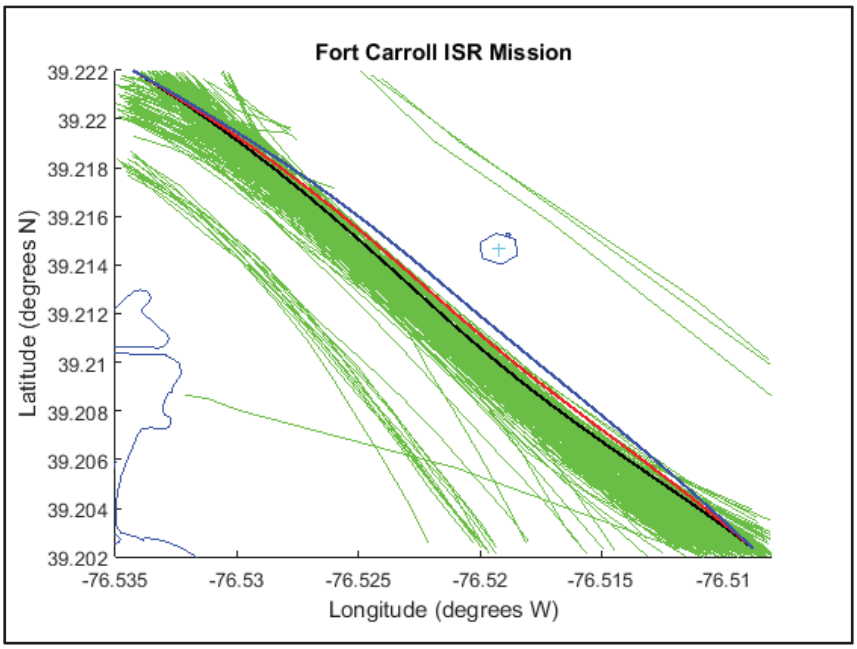

Figure 30: Fort Carroll ISR mission paths

VI. Validation of the Path's Optimality

The next step is validating the solution paths generated by DIDO. The solution to an optimal control problem is called an extremal. The extremal must satisfy three fundamental principles to be a correct and optimal solution: The Feasibility Principle, Bellman's Principle, and Pontryagin's Minimum Principle. The validation process described in this section is for the example problem introduced in Figure 27.

Path Feasibility

The first step in the path validation process is checking that the DIDO solution path is a feasible path for the USV to follow. This is performed by propagating DIDO's control trajectory through the system dynamics to obtain the state trajectory. This state trajectory is then compared to the DIDO calculated trajectory. If the two trajectories match, the DIDO solution is considered to be feasible.

The propagation of the vessel dynamics is performed in MATLAB's Simulink application. The Simulink diagram shown in Figure 31 is a visual representation of the USV's equations of motion expressed in Eq. (11).

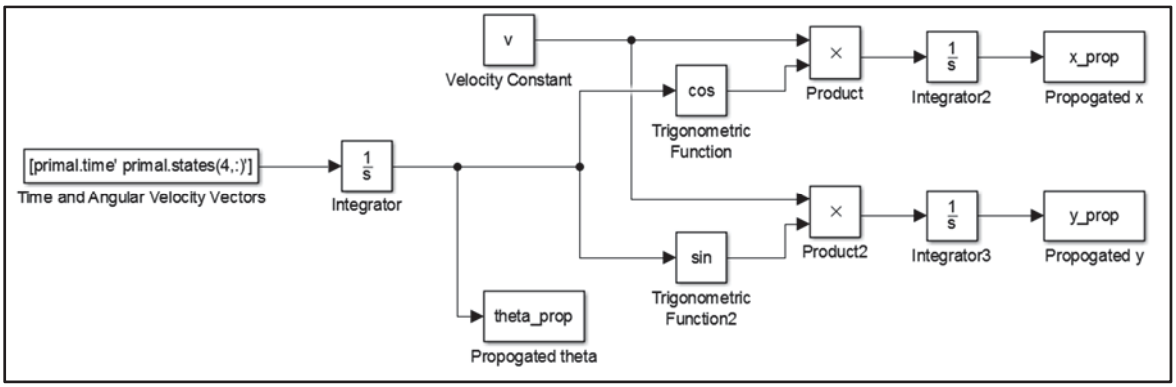

Figure 31: MATLAB Simulink model of USV

The vessel simulation is performed by using the DIDO calculated angular velocity values as the control input for the USV. The angular velocity values are inserted into the USV model according to the time vector calculated by DIDO. The DIDO software solves the path planning problem where the control is defined as the USV's angular acceleration. However, the calculated angular velocity is propagated in the feasibility calculations because its values are less erratic over time than the angular acceleration values. Figure 32 shows a comparison of the angular acceleration, shown in blue, and angular velocity, shown in red, of the USV in the sample problem introduced in Figure 27. The USV's angular acceleration has more extreme values, particularly near the beginning and end of the route, than the USV's angular velocity.

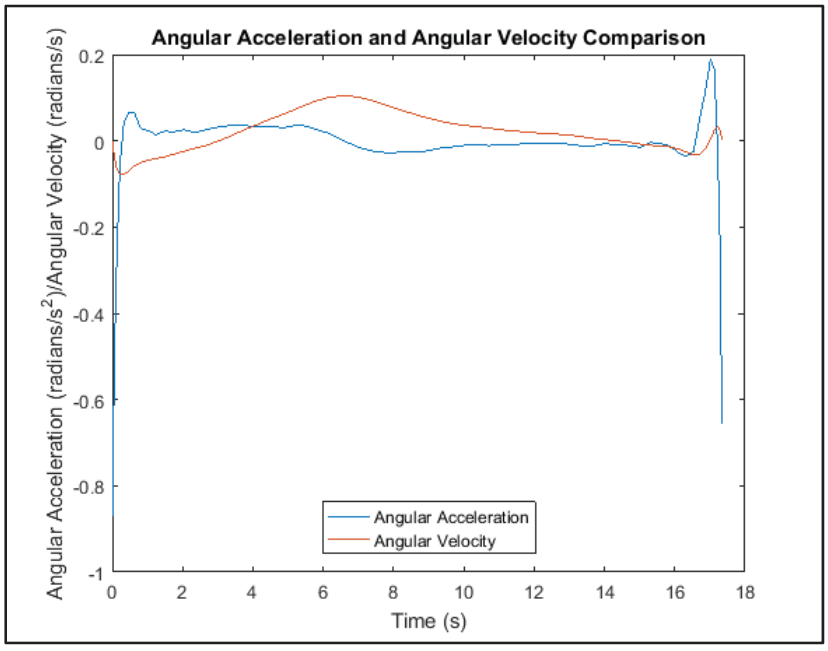

Figure 32: Comparison plot of angular acceleration and angular velocity values

Figure 33 shows the comparison for this same example problem of the solution trajectory computed by DIDO, indicated by the red line, and the trajectory calculated by propagating the angular velocity values in Simulink, marked by the blue asterisks. The coincidence of the two trajectories indicates that the DIDO solution is feasible.

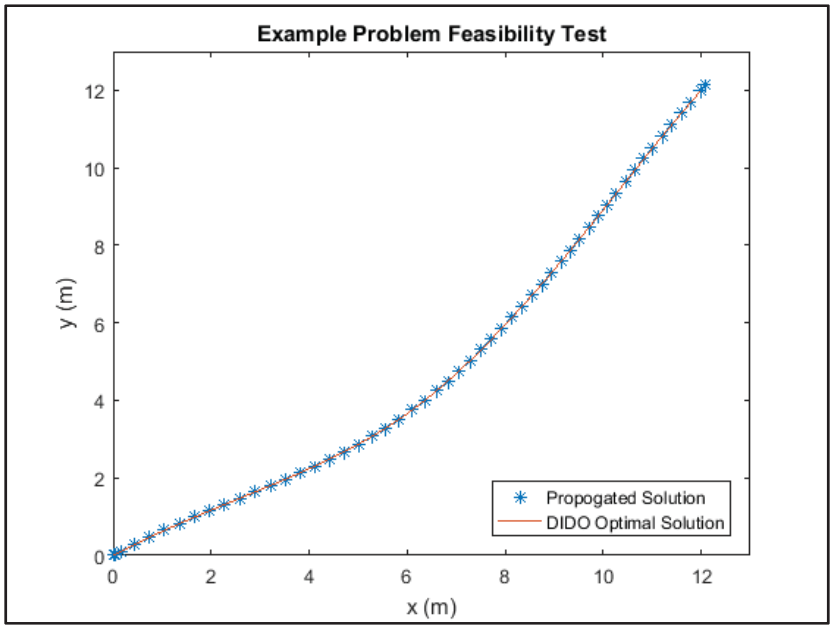

Figure 33: Feasibility test

Bellman's Principle

The next step in the path validation process is checking that the solution path satisfies Bellman's Principle. Bellman's Principle is illustrated in Figure 34 where an optimum path has been generated between points A and B and point C is a point located anywhere along this path. According to Bellman's Principle, if optimum paths are independently generated between A and C and between C and B, the concatenation of these two paths is equivalent to the optimum path from A to B. Furthermore, the cost of going from A to B is equivalent to the cost of going from A to C plus the cost of going from C to B [20].

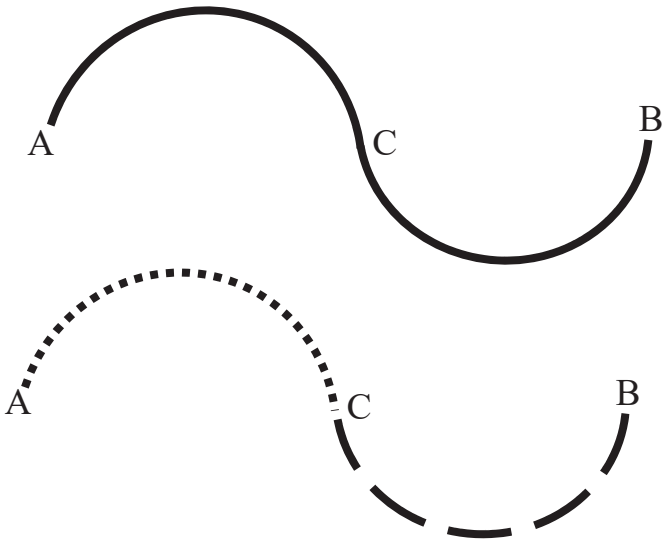

Figure 34: Bellman's Principle

Bellman's Principle is validated on the USV solution paths generated by DIDO by selecting a point anywhere between the starting point and the ending point on the full solution path. DIDO is then used to solve for the path going from the starting point to the states at the midpoint and again from the states at the midpoint to the ending point. Figure 35 shows all three solution paths for the example problem. The fact that that the first path segment and the second path segment are coincidental with the full length solution path proves the first requirement of Bellman's Principle. The second requirement of Bellman's Principle is that the cost associated with the total solution path is equal to the sum of the costs associated with the individual path segments. In the example below, the first path segment has a cost of 69.9 and the second segment has a cost of 174.0. These two costs sum to 243.9 which matches the cost of the full solution path. This confirms that the given extremal validates Bellman's Principle.

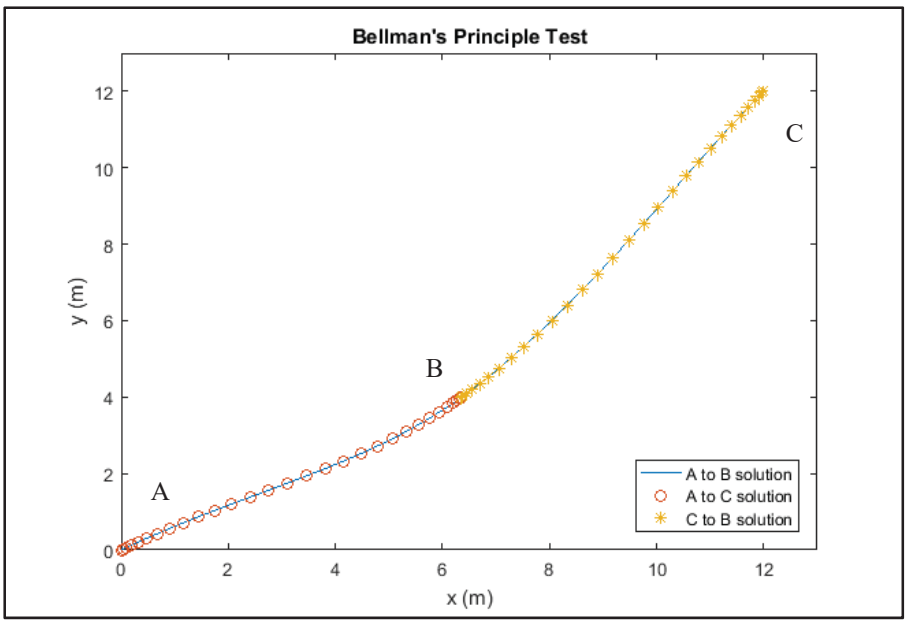

Figure 35: Bellman's Principle validation

Pontryagin's Minimum Principle

Proving that the extremal satisfies Pontryagin's Minimum Principle is the final step in validating that the extremal is in fact the optimal solution to the given control problem. This is accomplished by proving that the DIDO solution meets the five necessary conditions of optimality: the Hamiltonian Minimization Condition, the Adjoint Equation, the Transversality Condition, the Hamiltonian Value Condition, and the Hamiltonian Evolution Equation [21].

The Hamiltonian, H, is defined in Eq. (21) where $\underline{\lambda}$ is a Lagrange multiplier referred to as the costate.

$$H(\underline{\lambda}, \underline{x}, \underline{u}, t) = F(\underline{x}, \underline{u}, t) + \underline{\lambda}^T \underline{f}(\underline{x}, \underline{u}, t) \tag{21}$$

The Hamiltonian Minimization Condition (HMC) states that the Hamiltonian is to be minimized with respect to $\underline{u}$ while $\underline{\lambda}$ and $\underline{x}$ are held constant. The formal formulation of the HMC is shown in Eq. (22).

$$\begin{aligned} &\text{Minimize with respect to } \underline{u}\text{: } H(\underline{\lambda}, \underline{x}, \underline{u}, t) \\ &\text{Subject to: } \underline{h}^L \leq \underline{h}(\underline{x}(t), \underline{u}(t), t) \leq \underline{h}^U \end{aligned} \tag{22}$$

The path constraints must be incorporated into the HMC because the path constrains have the potential to be a function of $\underline{u}$. This can be accomplished by taking the Lagrangian of the Hamiltonian as shown in Eq. (23). The Lagrange multiplier associated with the path constraints is denoted by $\underline{\mu}$ and is named the state covector.

$$\overline{H}\left(\underline{\mu}, \underline{\lambda}, \underline{x}, \underline{u}, t\right) = H(\underline{\lambda}, \underline{x}, \underline{u}, t) + \underline{\mu}^T \underline{h}(\underline{x}, \underline{u}, t) \tag{23}$$

Karush-Kuhn-Tucker (KKT) conditions then must be introduced in order for the Lagrangian of the Hamiltonian to work with path constraints that have inequalities. The KKT conditions are shown in Eq. (24).

$$\mu_i \begin{cases} \leq 0 \quad if \quad h_i(\underline{x}, \underline{u}, t) = h_i^L \\ = 0 \quad if \quad h_i^L < h_i(u) < h_i^U \\ \geq 0 \quad if \quad h_i(\underline{x}, \underline{u}, t) = h_i^U \\ unrestricted \quad if \quad h_i^L = h_i^U \end{cases} \tag{24}$$

The Lagrangian of the Hamiltonian must be stationary with respect to the control as expressed in Eq. (25). The control solution must meet this requirement in order to satisfy the HMC.

$$\frac{\partial \overline{H}}{\partial u} = \underline{0} \tag{25}$$

The second component of Pontryagin's Minimum Principle is the Adjoint Equation. The Adjoint Equation expresses the relationship between the first derivative of the costates and the partial derivative of the Hamiltonian with respect to the states as shown in Eq. (26).

$$-\underline{\dot{\lambda}} = \frac{\partial H}{\partial \underline{x}} \tag{26}$$

It is next necessary to define the Endpoint Lagrangian with $\underline{\upsilon}$ serving as the Lagrange multiplier as expressed in Eq. (27).

$$\overline{E}(\underline{\upsilon}, \underline{x}_0, \underline{x}_f, t_0, t_f) = E(\underline{x}_0, \underline{x}_f, t_0, t_f) + \underline{\upsilon}^T \underline{e}(\underline{x}_0, \underline{x}_f, t_0, t_f) \tag{27}$$

This Endpoint Lagrangian is then used in the third component of Pontryagin's Minimum Principle, the Transversality Condition. The Transversality Condition requires the solution to satisfy the two relationships expressed in Eq. (28) between the costates and the partial derivative of the Endpoint Lagrangian with respect to the states.

$$\underline{\lambda}(t_f) = \frac{\partial \overline{E}}{\partial \underline{x}(t_f)}$$

$$\underline{\lambda}(t_0) = \frac{-\partial \overline{E}}{\partial \underline{x}(t_0)} \tag{28}$$

The fourth component of Pontryagin's Minimum Principle is the Hamiltonian Value Condition that is related to the Transversality Condition. The Hamiltonian Value Condition is expressed in two parts in Eq. (29) as a relationship between the Hamiltonian and the Endpoint Lagrangian.

(29)

$$H(t_f) = \frac{-\partial \overline{E}}{\partial t_f}$$

$$H(t_0) = \frac{\partial \overline{E}}{\partial t_0}$$

The final component of Pontryagin's Minimum Principle is the Hamiltonian Evolution Equation. Pontryagin's *Minimum* Principle solves for the Lower Hamiltonian. The optimal solution to the control problem will exist where the Hamiltonian of the optimal control solution, $\underline{u}^*$, is at a minimum. This minimum, expressed as $\mathcal{H} = H(\underline{u}^*)$, is named the Lower Hamiltonian. The Hamiltonian Evolution Equation, expressed in Eq. (30), describes how the change of the Lower Hamiltonian over time is related to the Lagrangian of the Hamiltonian.

$$\dot{\mathcal{H}} = \frac{\partial \overline{H}}{\partial t} \tag{30}$$

The optimal solution generated by DIDO can be validated with Pontryagin's Minimum Principle with the outputs provided by DIDO. DIDO provides the states, costates, covectors, controls, Hamiltonian, and other outputs necessary to verify that the optimal solution satisfies all of Pontryagin's components. The figures and calculations below provide an example of how some of the Pontryagin components can be validated for the sample problem introduced in Figure 27. The example problem introduced above is formally defined in Eq. (31).

$$\text{Given: } \underline{x} = \begin{bmatrix} x \\ y \\ \theta \\ \omega \end{bmatrix} \in X \subset \mathbb{R}^4 \qquad \underline{u} = [\alpha] \in \mathbb{R} \tag{31}$$

$$\text{Find: } \underline{x}(\cdot), \underline{u}(\cdot)$$

$$\text{That Minimizes: } J = \int_{t_0}^{t_f} \left[(y - y_m)^2 + W[(x - x_T)^2 + (y - y_T)^2] \right]$$

$$\text{Subject to: } \dot{\underline{x}} = \begin{bmatrix} \dot{x} \\ \dot{y} \\ \dot{\theta} \\ \dot{\omega} \end{bmatrix} = \begin{bmatrix} v\cos\theta \\ v\sin\theta \\ \omega \\ \alpha \end{bmatrix}$$

$$0 \le x \le 12$$
$$0 \le y \le 12$$
$$\frac{-\pi}{2} \le \theta \le \pi$$
$$-1 \le \omega \le 1$$
$$-1 \le \alpha \le 1$$
$$v = 1$$
$$W = 0.2$$

$$\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 12 \\ 12 \\ 0 \\ 0 \end{bmatrix} \leq \underline{e}(\underline{x}_0, \underline{x}_f, t_0, t_f) = \begin{bmatrix} t_0 \\ x(t_0) \\ y(t_0) \\ \theta(t_0) \\ \omega(t_0) \\ x(t_f) \\ y(t_f) \\ \theta(t_f) \\ \omega(t_f) \end{bmatrix} \leq \begin{bmatrix} 0 \\ 0 \\ 0 \\ \frac{\pi}{2} \\ 0 \\ 12 \\ 12 \\ \frac{\pi}{2} \\ 0 \end{bmatrix}$$

The next step is calculating the Hamiltonian as shown in Eq. (32).

$$H = (y - y_m)^2 + W[(x - x_T)^2 + (y - y_T)^2] + \lambda_x v \cos\theta + \lambda_y v \sin\theta + \lambda_\theta \omega + \lambda_\omega \alpha \quad (32)$$

The KKT conditions are shown in Eq. (33).

$$\mu_x \begin{cases} \leq 0 \quad if \quad x = 0 \\ = 0 \quad if \quad 0 < x < 12 \\ \geq 0 \quad if \quad x = 12 \end{cases} \quad (33)$$

$$\mu_y \begin{cases} \leq 0 \quad if \quad y = 0 \\ = 0 \quad if \quad 0 < y < 12 \\ \geq 0 \quad if \quad y = 12 \end{cases}$$

$$\mu_\theta \begin{cases} \leq 0 \quad if \quad \theta = 0 \\ = 0 \quad if \quad \frac{-\pi}{2} < \theta < \pi \\ \geq 0 \quad if \quad \theta = \pi \end{cases}$$

$$\mu_\omega \begin{cases} \leq 0 \quad if \quad \omega = -1 \\ = 0 \quad if \quad -1 < \theta < 1 \\ \geq 0 \quad if \quad \omega = 1 \end{cases}$$

$$\mu_\alpha \begin{cases} \leq 0 \quad if \quad \alpha = -1 \\ = 0 \quad if \quad -1 < \alpha < 1 \\ \geq 0 \quad if \quad \alpha = 1 \end{cases}$$

The Lagrangian of the Hamiltonian is expressed in Eq. (34).

$$\overline{H} = (y - y_m)^2 + W[(x - x_T)^2 + (y - y_T)^2] + \lambda_x v \cos\theta + \lambda_y v \sin\theta + \lambda_\theta \omega + \lambda_\omega \alpha + \mu_x x + \mu_y y + \mu_\theta \theta + \mu_\omega \omega + \mu_\alpha \alpha \quad (34)$$

HMC is then applied as shown in Eq. (35).

$$\frac{\partial \overline{H}}{\partial \alpha} = \lambda_\omega + \mu_\alpha = 0 \quad \rightarrow \quad \lambda_\omega = -\mu_\alpha \tag{35}$$

The HMC for this problem is validated by Figure 36 which shows that $\lambda_\omega = -\mu_\alpha$ throughout the duration of the problem. A numerical comparison of the values can also be performed.

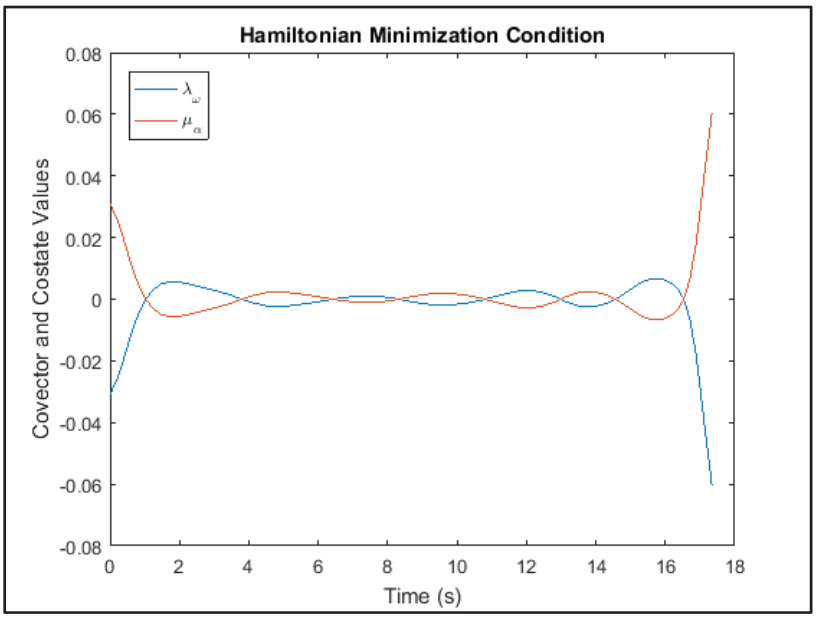

Figure 36: Hamiltonian Minimization Condition validation

Solving the Adjoint Equations results in Eq. (36).

$$\begin{aligned} \dot{\lambda}_x &= -2W(x - x_T) - \mu_x \\ \dot{\lambda}_y &= -2(y - y_m) - 2W(y - y_T) - \mu_y \\ \dot{\lambda}_\theta &= \lambda_x v \sin\theta - \lambda_y v \cos\theta - \mu_\theta \\ \dot{\lambda}_\omega &= -\lambda_\theta - \mu_\omega \end{aligned} \tag{36}$$

Figure 37 shows an example of how to validate the fourth statement in Eq. (38), $\dot{\lambda}_\omega = -\lambda_\theta - \mu_\omega$. The blue line in the image shows $(-\lambda_\theta - \mu_\omega)$ and the orange line shows λ_ω. Visual inspection confirms that the slope of λ_ω is positive when $(-\lambda_\theta - \mu_\omega)$ is positive in value and the slope of λ_ω is negative when $(-\lambda_\theta - \mu_\omega)$ is negative. These observations visually confirm that the given solution is consistent with the Adjoint Equation component of Pontryagin's Minimum Principle.

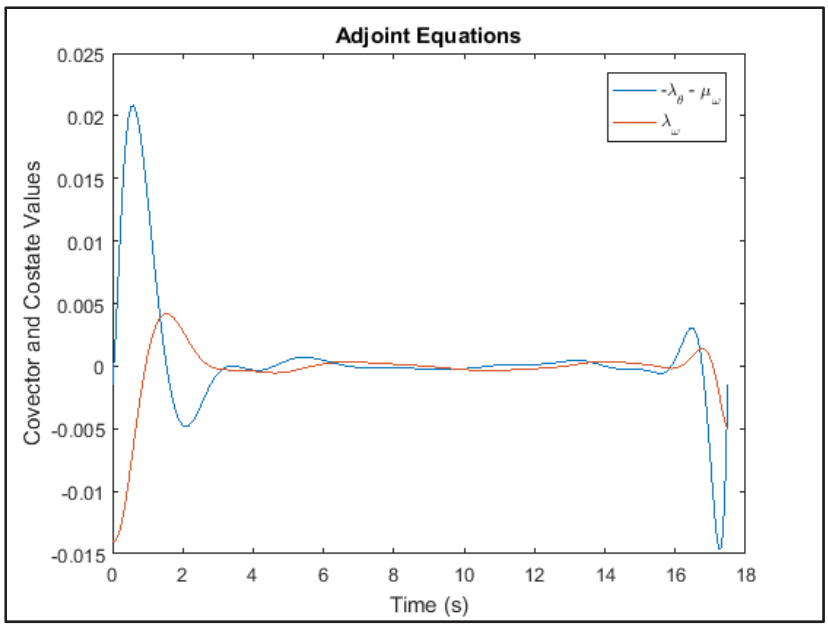

Figure 37: Adjoint Equation validation

The Endpoint Lagrangian is calculated as shown in Eq. (37).

$$\overline{E} = \upsilon_{t_0} t_0 + \upsilon_{x_0} x(t_0) + \upsilon_{y_0} y(t_0) + \upsilon_{\theta_0} \theta(t_0) + \upsilon_{\omega_0} \omega(t_0) + \upsilon_{x_f} x(t_f) + \upsilon_{y_f} y(t_f) + \upsilon_{\theta_f} \theta(t_f) + \upsilon_{\omega_f} \omega(t_f) \tag{37}$$

The KKT conditions for the endpoint constraints where $\underline{e}^L \neq \underline{e}^U$ are shown in Eq. (38).

$$\upsilon_{\theta_0} = \begin{cases} \leq 0 & if \quad \theta(t_0) = 0 \\ 0 & if \quad 0 < \theta(t_0) < \dfrac{\pi}{2} \\ \geq 0 & if \quad \theta(t_0) = \dfrac{\pi}{2} \end{cases} \tag{38}$$

$$\upsilon_{\theta_f} = \begin{cases} \leq 0 & if \quad \theta(t_f) = 0 \\ 0 & if \quad 0 < \theta(t_f) < \dfrac{\pi}{2} \\ \geq 0 & if \quad \theta(t_f) = \dfrac{\pi}{2} \end{cases}$$

The resulting Transversality Conditions are shown in Eq. (39).

$$\begin{aligned} \lambda_x(t_0) &= -v_{x_0} \\ \lambda_y(t_0) &= -v_{y_0} \\ \lambda_\theta(t_0) &= -v_{\theta_0} \\ \lambda_\omega(t_0) &= -v_{\omega_0} \\ \lambda_x(t_f) &= v_{x_f} \\ \lambda_y(t_f) &= v_{y_f} \\ \lambda_\theta(t_f) &= v_{\theta_f} \\ \lambda_\omega(t_f) &= v_{\omega_f} \end{aligned} \tag{39}$$

The costates are plotted over time in Figure 38. Equation (40) expresses the values for the endpoint Lagrange multipliers. A comparison of Figure 38 and Equation (40) confirms that all of the Transversality Conditions are met. The initial costates, $\lambda_i(t_0)$, are equal to the negative of the initial endpoint Lagrange multipliers, v_{i_0}. The final costates are equal to the final endpoint Lagrange multipliers.

$$\begin{bmatrix} v_{x_0} \\ v_{y_0} \\ v_{\theta_0} \\ v_{\omega_0} \\ v_{x_f} \\ v_{y_f} \\ v_{\theta_f} \\ v_{\omega_f} \end{bmatrix} = \begin{bmatrix} 8.90 \\ 4.56 \\ 0.00 \\ 0.01 \\ -10.47 \\ -16.57 \\ 0 \\ -0.01 \end{bmatrix} \tag{40}$$

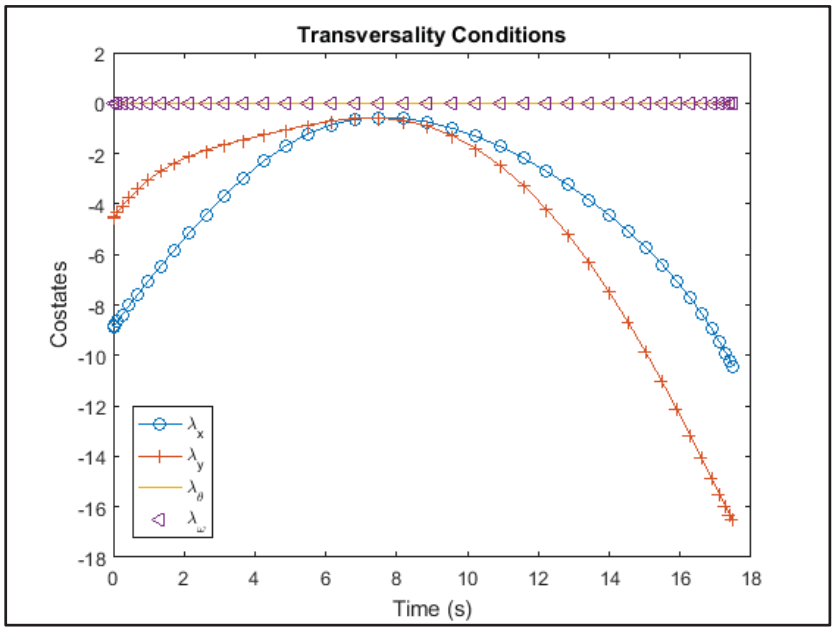

Figure 38: Transversality Condition validation

The Hamiltonian Value Condition is calculated in Eq. (41) and the Hamiltonian Evolution Equation is calculated in Eq. (42).

$$H(t_0) = \upsilon_{t_0} \qquad (41)$$
$$H(t_f) = 0$$

$$\dot{H} = 0 \qquad (42)$$

These conditions together indicate that the Hamiltonian is equal to 0 throughout the solution. Figure 39 confirms that the Hamiltonian remains approximately equal to 0 for the entire length of the path.

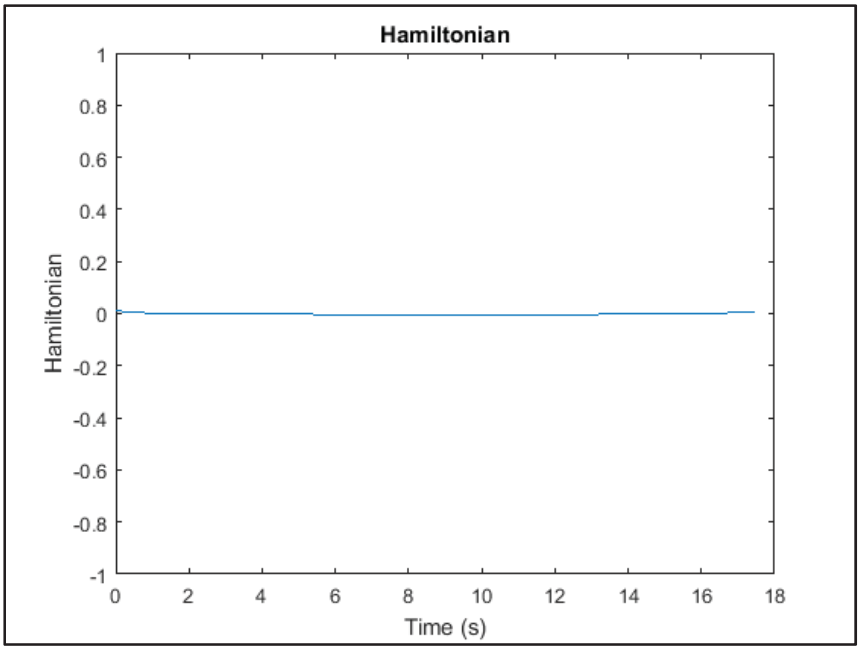

Figure 39: Hamiltonian Value Condition and Hamiltonian Evolution Equation validation

Pontryagin's Minimum Principle is satisfied when all five necessary conditions for optimality are met. The optimal solution is validated once it is proven to satisfiy the Feasibility Principle, Bellman's Principle, and Pontryagin's Minimum Principle.

VII. Evaluation of the Optimal Path's Identifiability and Quality of Intelligence

Once the DIDO generated solution has been validated, the next step is to evaluate the path's ability to allow the USV to accomplish its mission objectives. This step determines how desirable the generated path is for a USV to follow during an ISR mission. The path is evaluated on the two principal mission objectives of gathering intelligence on the target of interest and avoiding identification by the adversary.

Quality of Intelligence

The quality of the intelligence gathered by the USV depends on the distance between the USV and the target and the length of time that the USV is in the vicinity of the target. The distance between a vessel's track and the target is measured with the framework of the target rays approach. As introduced above, the target rays approach measures the distance between a track and the target along a series of rays radiating from the target. This approach is introduced originally as a possible means of determining which track segments to select in each sector surrounding the target. It is used in the evaluation section as a means of measuring how successful the USV is at drawing close to the target.

The average distance between a track and the target is calculated with Eq. (2). As a part of the evaluation section, the target rays approach is calculated with all of the rays combined into one large sector. The number of rays per sector, N_{rs}, is therefore equal to the total number of rays.

$$\overline{D}_{i,k} = \frac{\sum_{j=1}^{N_{rs}} d_{i,j,k}}{N_{rs}} \tag{2}$$

In the Fort Carroll example problem, a USV's distance from Fort Carroll is measured along 29 consecutive rays. Figure 40 provides a visual of the target rays, shown as dashed black lines, observed vessel tracks, shown in green, and a generated USV path, shown as the solid black line.

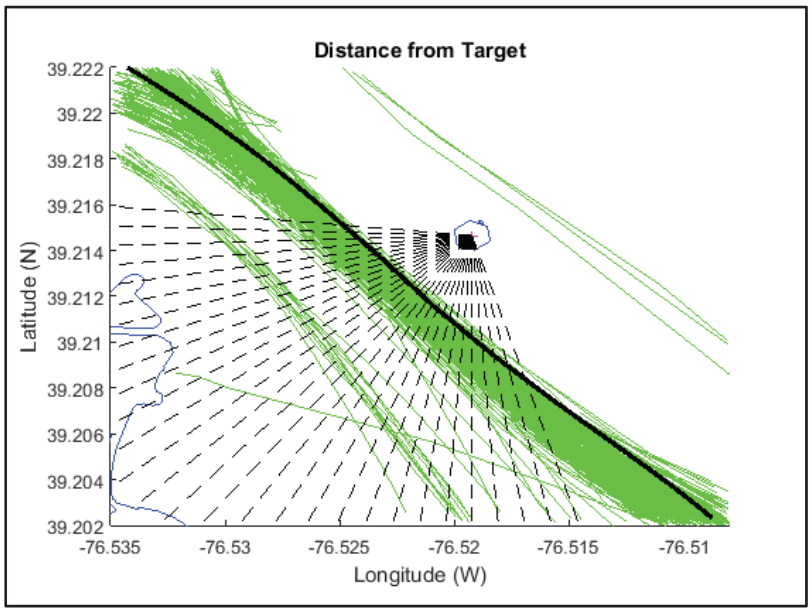

Figure 40: Evaluation of average distance from track to target

The actual value of $\overline{D}_{i,k}$ for an optimal track is not as valuable as its relative magnitude compared to the $\overline{D}_{i,k}$ of other tracks. It is therefore beneficial to determine the percentile of the $\overline{D}_{i,k}$ of an optimal track in relation to the $\overline{D}_{i,k}$ values of the observed vessel tracks. Figure 41 shows the distribution of $\overline{D}_{i,k}$ for the observed tracks in the Fort Carroll problem. The $\overline{D}_{i,k}$ from a generated optimal track is compared to this distribution to determine the percentage of observed tracks of which the optimal track has a lower average distance to the target.

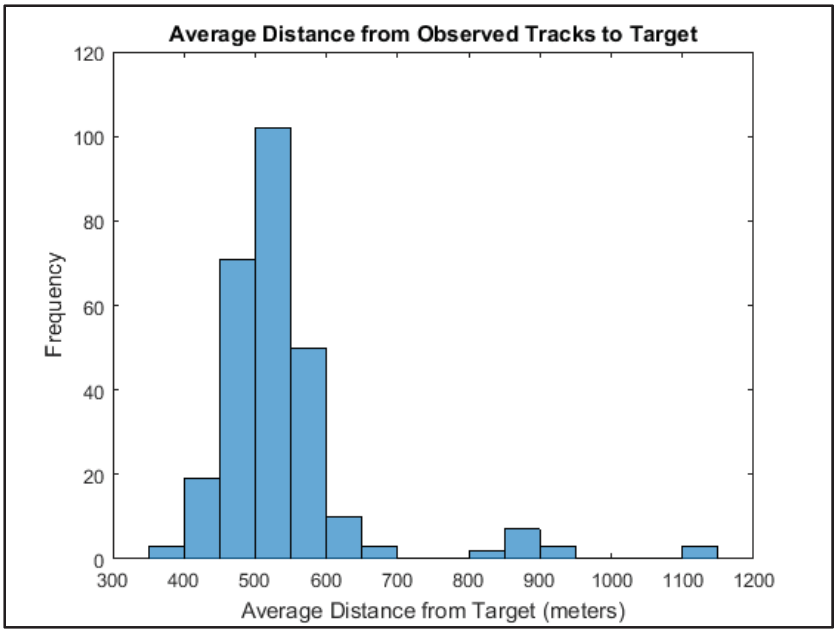

Figure 41: Observed track average distances

The second metric related to the quality of intelligence gathered by the USV is the length of time that the USV spends in the vicinity of the target. The precise calculation of this value involves measuring the time elapsed from when the USV enters a particular range of the target and departs that range. However, this precision is not necessary because of the fact that most vessels follow the same general path through the area of interest. The length of time that a vessel spends in the vicinity of the target is therefore related to the vessel's average speed through the area around the target. In Section IV it is discussed that the USV's speed is determined by selecting the speed that exists at a given percentile of the observed vessel speeds. A lower percentile corresponds to having a lower speed and therefore a longer period of time in the vicinity of the target. The USV's average speed percentile is used in the evaluation section as the metric related to the length of time that the USV spends in the vicinity of the target.

The metric for the time in the vicinity of the target is less useful than the metric for the distance between the target and the USV. This research focuses more on the USV's location within the area of interest than how fast the USV travels through the given area. Therefore, this research does not focus on evaluating USV tracks at a wide range of speeds.

Risk of Identification

The next step is to evaluate the generated optimal path to determine whether the adversary would identify the USV as a suspicious vessel if it were to follow the path during an ISR mission. There has already been significant research performed into how to detect anomalies in vessel traffic. At a workshop of Swedish maritime professionals in 2008, experts came up with 75 unique anomalies in maritime traffic spanning categories of cargo, suspicious crew, and vessel movements [22]. Because of the utility and availability of AIS data, many of these techniques have been designed specifically to analyze AIS transmissions. Some techniques involve comparing the information in the AIS messages to the actions of the vessel. For example, a vessel is labeled as an anomaly if it is not heading towards its stated destination. Other anomalies

of this nature involve checking if registration details about a vessel, such as the vessel's name, change between AIS transmissions or if the vessel turns its AIS transmitter off and on [23]. However, this research strictly considers a USV's location and movements when determining whether the USV appears to be an anomaly. Addressing the other anomalies is beyond the scope of this research.

The adversary is assumed to have anomaly detection software and sensors capable of observing the USV along its entire route. If the USV is conducting an operation in a hostile environment where the normal maritime traffic uses AIS, it is advantageous for the USV to use AIS as well so that it does not draw suspicion from onlookers in being a vessel that should have AIS but is not transmitting its location. In this instance, the adversary is assumed to be analyzing the USV's AIS transmissions. However, the USV is also expected to operate in locations where the normal maritime traffic does not use AIS or it is more advantageous for the USV to not transmit with AIS. In these instances, the adversary is assumed to be monitoring the USV's location with RADAR or electro-optical sensors. It is therefore assumed that the adversary is monitoring the USV's location at all times.

Most anomaly detection techniques identified by researchers focus on generating Probability Density Functions (PDFs) of the latitudinal and longitudinal positions and velocities of observed vessels. Many of these techniques involve using self-learning computer programs and large data sets in order to generate the PDFs with minimal input from the researcher. Two means of creating these distributions, with a Gaussian Mixture Model (GMM) or Kernel Density Estimator (KDE), have been supported with significant research. Laxhammer and others [24] indicated that both of these methods were equally effective at discriminating anomalous traffic from normal maritime traffic, but discrimination took longer than the authors hypothesized would be necessary. Pallotta and others built on this work by creating a self-learning program named Traffic Route Extraction and Anomaly Detection (TREAD) [25]. When applied to a given region, the TREAD approach identifies exit points, entry points, and stationary points in the area based on the vessel data. Route objects are then created between these identified waypoints. The distribution of velocity vectors along these routes are recorded. The velocity vectors of new vessels are added to these distributions as determined by a KDE. The TREAD methodology has proven to be effective at identifying anomalous vessel traffic.

While the TREAD approach is effective, the sophistication of its design makes it complicated and time-consuming to implement. TREAD requires self-learning code capable of generating waypoints, routes, and velocity vector distributions from large data sets. Pseudo-code for this approach is available in [25]. Another shortcoming of TREAD is that it is designed without any particular target or suspicious behavior in mind. This means that it may not have the sensitivity to detect anomalous behavior in the vicinity of the designated target at the desired level. Because of the complexity of TREAD and concerns over its effectiveness for this application, TREAD is not used for this research.

The anomaly detection process used in this research focuses on comparing the USV's path to the collection of observed vessel paths for the same given area. There are a number of parameters that can be used to compare the USV's planned path to the observed paths. Some of these parameters are the distance between a path and the target along a particular ray, the average

distance between a path and the target, the average speed of a vessel transiting along the path, and the variation in the vessel's speed along the route. This research focuses on the following two parameters: the distance between a path and the calculated mean path and a measure of how much the USV moves laterally within the observed vessel traffic. These two parameters are important in determining whether the USV is able to appear as a normal vessel as it follows the generated path. Analysis is not performed on the speed of the USV along the path because the USV's speed is selected from the distribution of observed vessel speeds and is held constant for the duration of the route. The anomaly detection approach used in this research is conducted with a particular target and set of suspicious behaviors in mind. For example, the adversary is likely to be more concerned about a vessel that is closer to the target than normal than the adversary is about a vessel that is twice as far from the target than normal.

The first parameter to be tested is the distance between a given track and the mean path at a set location along the route. This test is to determine whether the USV moves too far to the left or right of the observed vessel traffic. This parameter is most important in the portion of the route where the USV passes by the target. If the USV deviates too much from the observed vessel traffic to draw near the target, then the USV is liable to appear as a suspicious vessel to the adversary. This test is performed at numerous points along the mean path at a set interval. The random variable X_D^i represents the distance between the tracks and the mean path at a test point i. In the Fort Carroll problem, the index i goes from 1 to 21. At each test point, the distance between the vessel tracks and the mean path are calculated along perpendicular lines. Figure 42 shows one of these perpendicular lines, at test point #10, for the Fort Carroll example problem. The only observed vessel tracks of interest are those that are within the main channel. Therefore, distances are only calculated between these tracks and the mean path as shown by the length of the perpendicular line in Figure 42.

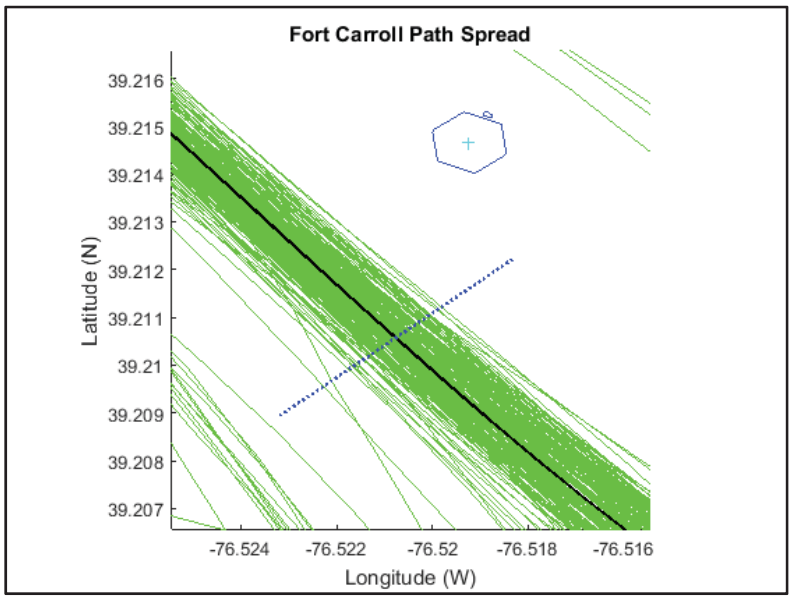

Figure 42: Spread of tracks along a line (dashed blue) perpendicular to the mean path (solid black)

The distribution of X_D^{10} for the observed vessel traffic at the test point denoted above is shown in Figure 43. X_D^{10} is represented by all northwesterly travelling vessel tracks within 140 meters of the mean path. The histogram in Figure 43 therefore shows the spread around the mean path of the entire population of observed vessels travelling in the northwest direction in the given channel for this month of data. Positive values represent tracks that are to the southwest of the mean path and are therefore further from Fort Carroll while negative values correspond with tracks that are northeast of the mean path and are closer to Fort Carroll.

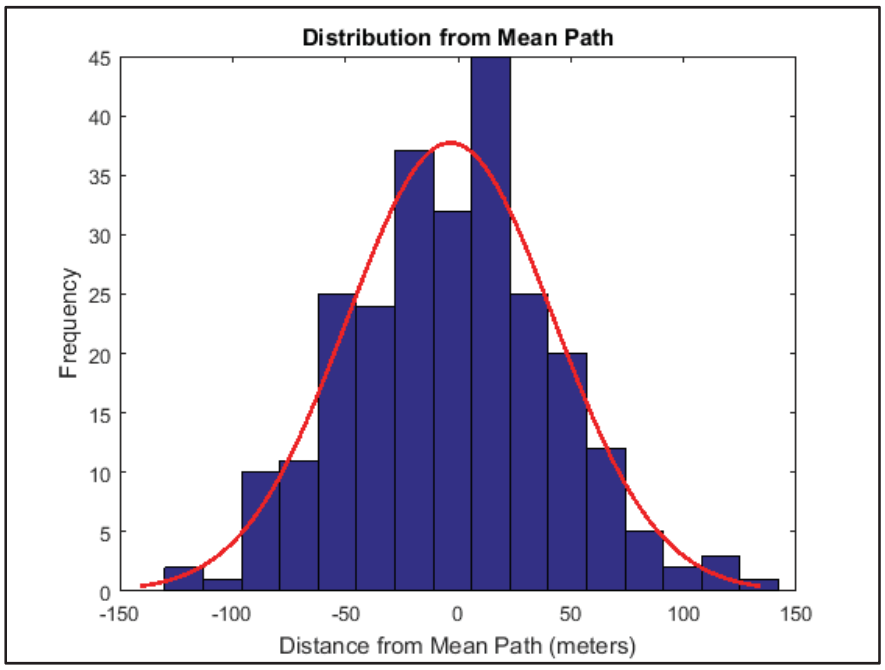

Figure 43: Distribution of X_D^{10}

The distribution of X_D^i at each test point for the observed traffic can be approximated with a Normal distribution. Because there are a large number of tracks and all of the northwest travelling vessel tracks are used, the mean and the standard deviation of X_D^i at each test point i are considered to be reasonable estimates of the population mean μ_i and standard deviation σ_i. Figure 44 shows the Normal PDF of X_D^{10}.

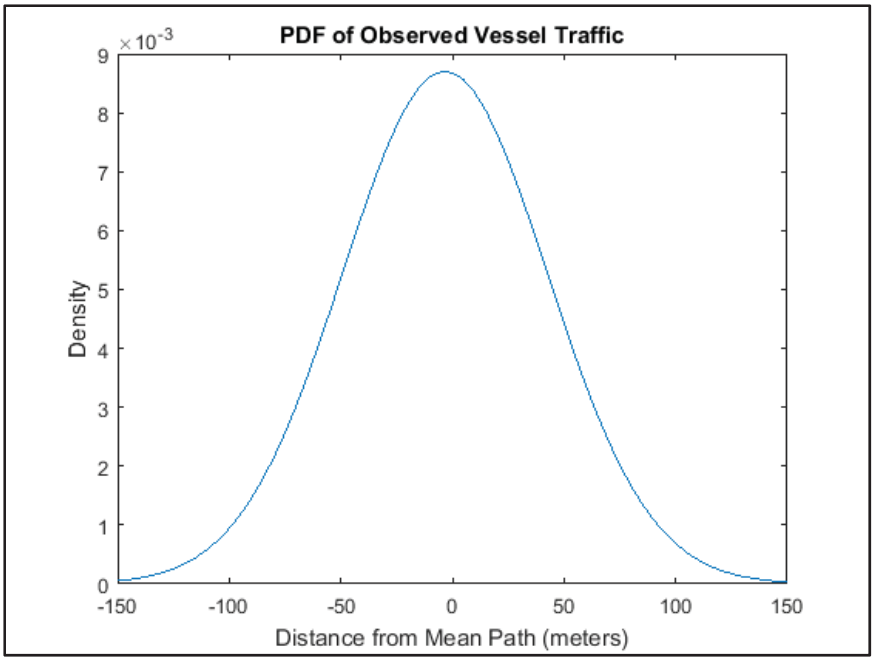

Figure 44: PDF of observed traffic

As discussed above, this research performs anomaly detection with a set of suspicious behaviors in mind. The adversary is assumed to know what sorts of behaviors are liable to be exhibited by a vessel conducting ISR operations. The behavior of the USV is compared to the behaviors exhibited by the observed vessel traffic and to the behaviors of the known ISR vessels. The behavior of the observed traffic is determined in the manner discussed above. However, it is more difficult to compare the USV's planned path to ISR vessel tracks because there is not a data set of ISR vessel tracks like there is a data set of normal maritime vessel tracks. It is therefore necessary to develop an approximation of the tracks that ISR vessels would follow if they were to gather intelligence on the same target of interest on which the USV is targeting.

A set of USV path solutions for a given ISR mission is used as the approximation of the tracks used by ISR vessels on that given target. In essence, the paths developed by this research are used to approximate the behavior of ISR vessels. A set of USV path solutions is generated by varying the weight W in the path planning cost function given in Eq. (19). Figure 45 shows a set of 21 USV solution paths overlaid on the observed vessel tracks. Figure 46 provides an inset of the tracks in the area closest to Fort Carroll.

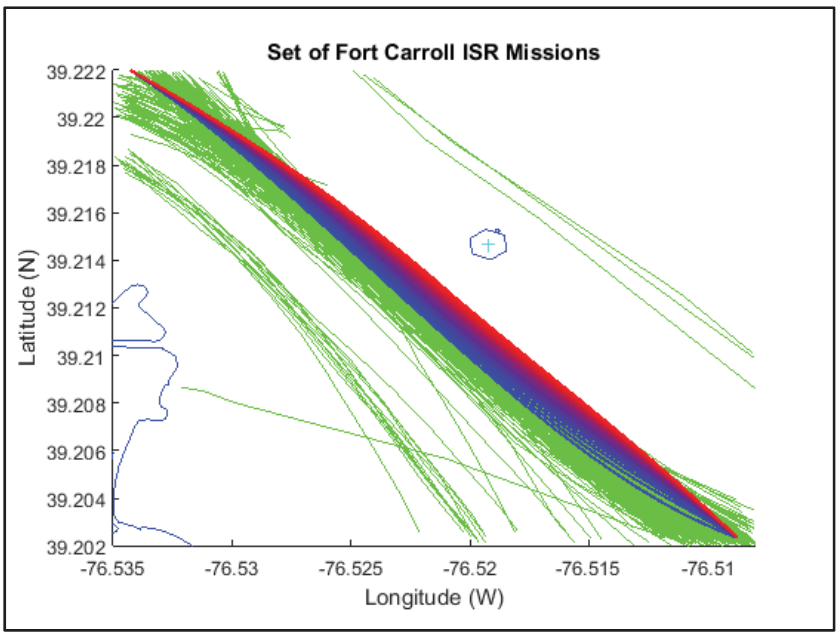

Figure 45: Fort Carroll mission tracks

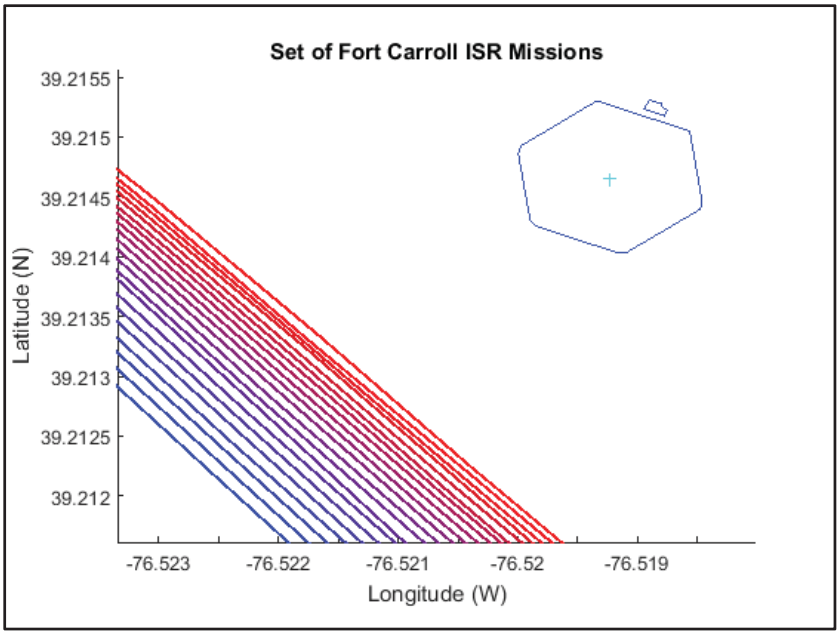

Figure 46: Fort Carroll mission tracks inset

The distribution of X_D^i for the ISR traffic is approximated by this set of 21 tracks. The ISR distribution of X_D^{10} in the Fort Carroll problem is shown in Figure 47. Because the ISR tracks are developed with Fort Carroll as the target, all of the ISR tracks are closer to Fort Carroll than the mean path.

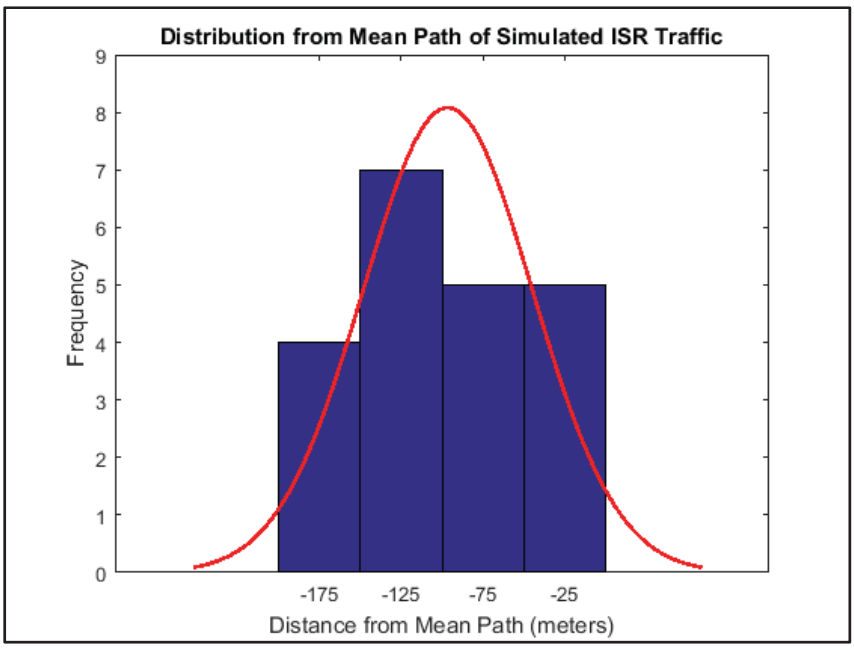

Figure 47: Distribution of X_D^i at test point for simulated ISR traffic

The distribution of X_D^i for the ISR traffic is also approximated with a Normal distribution. The Normal distribution approximation of the ISR traffic is not as strong as the Normal distribution approximation of the observed vessel traffic because the sample of ISR tracks is smaller and the ISR tracks are not selected in a random manner. However, Figure 47 indicates that the distribution of X_D^{10} for the ISR traffic is unimodal and a Normal distribution provides a reasonable approximation. A comparison of the PDFs for X_D^{10} for the observed and simulated ISR traffic is shown in Figure 48. As anticipated, the mean of the ISR traffic, -96.5 meters, is less than the mean of the observed traffic, -3.3 meters, indicating that the ISR tracks are closer to Fort Carroll. The standard deviation of the ISR distribution, 51.8 meters, is greater than the standard deviation of the observed distribution, 45.8 meters, indicating that the ISR tracks tend to be more dispersed than the observed tracks.

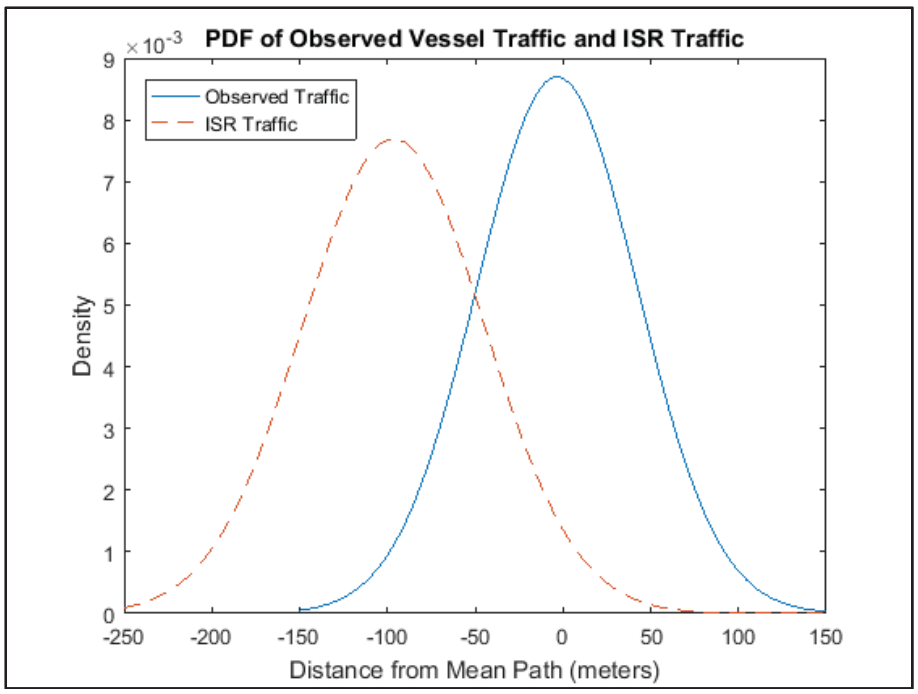

Figure 48: PDFs of observed and ISR traffic

Anomaly detection in regards to X_D^i is performed via hypothesis testing. This involves analyzing the generated optimal path for the USV and determining whether the adversary would believe that the USV is a member of the normal maritime traffic or if the USV is a suspicious vessel. The adversary's null hypothesis, H_0, is that the USV is an innocent vessel that is a member of the normal maritime traffic. The alternative hypothesis, H_1, is that the USV is not from the distribution of normal maritime traffic and may be a vessel conducting ISR. The adversary sets a threshold, designated as gamma (γ), at which it rejects the H_0 and accepts the H_1. Gamma is the distance from the mean path at which a vessel is first identified as being suspicious in nature. As the anomaly detection is being performed in relation to the target, one-sided hypothesis testing is used as the adversary is concerned with vessels that are closer to the target than the observed traffic.

In determining where to set gamma, the adversary is concerned with the two types of errors that can occur with hypothesis testing [26]. Type I error occurs when the adversary rejects the H_0 when the H_0 is correct. This error, referred to as the "significance level" of the test and expressed as α in Eq. (43), exists when a vessel is further from the mean path than γ but is actually an innocent member of the normal traffic. This error is undesirable because it corresponds to unnecessary intercepts by the adversary's security team and may result in angering the local traffic. The adversary will set a low α value in an attempt to minimize the number of unnecessary vessel interdictions. Type II error, Eq. (44), exists when the adversary fails to reject the H_0 when the H_1 is true. This occurs when a vessel is closer to the mean path than γ but is actually conducting ISR operations on the target. This error is very undesirable because it represents the probability that the adversary fails to identify an intelligence gathering vessel.

$$P(type\ I\ error) = \alpha = P(X_D^i \leq \gamma | H_0) \tag{43}$$

$$P(type\ II\ error) = P\left(X_D^i > \gamma \middle| H_1\right) \tag{44}$$

In the anomaly detection in this research, the simulated adversary takes its accepted levels of type I and type II error into account in determining where to set its threshold γ. The adversary must balance its accepted number of false-alarms, type I error, with how many ISR vessels it is willing to fail to identify, type II error. In order to represent a variety of mission scenarios, the adversary is assumed to have three different security states. In a heightened security state, the adversary is more aggressive at stopping vessels and is willing to intercept more innocent members of the normal maritime traffic. In a relaxed security state, the adversary only intercepts vessels that deviate significantly from the mean path towards the target and therefore has a lower false alarm rate. The normal security state exists in the middle. The security states are distinguished by the false alarm rates indicated in Table 2. At each test point, three hypothesis tests, one at each security state, are conducted on X_D^i. Although the false alarm rates are set at the same levels for each test point, each test point produces different type II errors for each security state because each test point has a different PDF for X_D^i. Table 2 shows the type II error and the detection rate, defined as $(1 - type\ II\ error)$, for the hypothesis tests conducted at test point #10 in the Fort Carroll problem. The table indicates that a relaxed security posture fails to identify about 60.3% of the ISR traffic. On the other hand, the heightened security posture is able to correctly identify 73.6% of the ISR traffic. The cost of this increased detection rate is the fact that about 10% of the innocent traffic is falsely intercepted in the heightened security state.

Table 2: Adversary security states at test point #10

Security State	False Alarm Rate (α)	Type II error	Detection Rate
Relaxed Security	1%	60.3%	39.7%
Normal Security	5%	36.6%	63.4%
Heightened Security	10%	25.4%	74.6%

Figure 49 depicts the gamma values that correspond to each security state in Table 2.

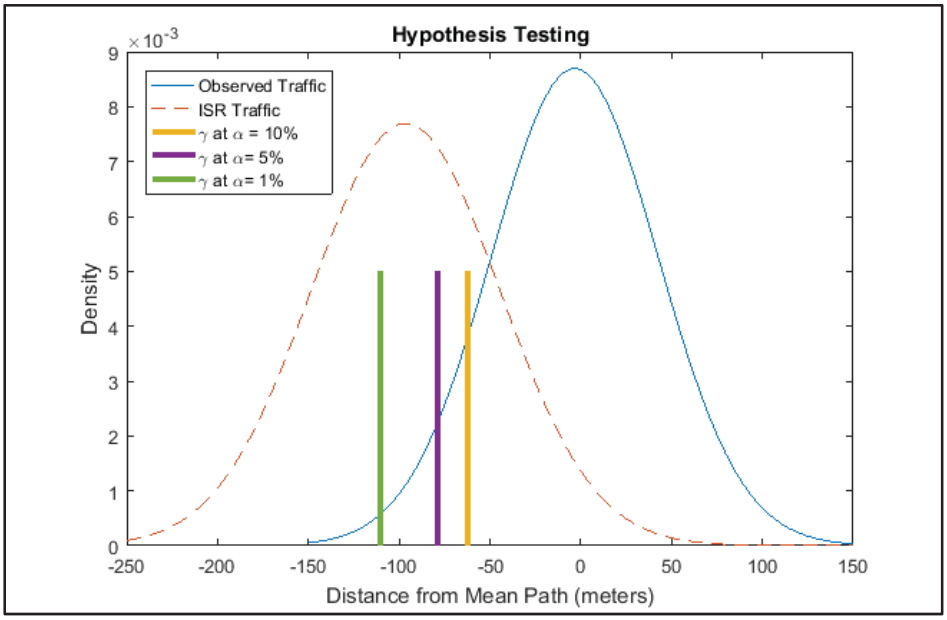

Figure 49: Hypothesis testing at test point #10

The second parameter tested as a part of the anomaly detection process is a measure of how much the USV's relative position within the channel of normal vessel traffic changes over the duration of its route. This parameter is intended to measure how much the USV "swerves" within the main body of the traffic. A USV is liable to draw the suspicion of the adversary if it makes s-turns through the main channel travelling from one side of the channel to the other throughout its route. The second parameter, X_S, measures the standard deviation of the relative position of a track within the vessel traffic for the duration of the route.

The first step in calculating X_S is normalizing the distributions of X_D^i at each test point. The distance at each test point between each track and the mean path is normalized by calculating the z-value as shown in Eq. (45). In this equation, x_i^t denotes the distance between an individual track t and the mean path at test point i. Each z_i^t is calculated with this value and the mean and standard deviation of the distribution at that point.

$$z_i^t = \frac{x_i^t - \mu_i}{\sigma_i} \tag{45}$$

The random variable X_S for a given track is equal to the standard deviation of the z-values at each test point. Each track has a z-value at each test point. Taking the standard deviation of this collection of z-values for each track allows one to analyze how much a track's relative position changes along the route. This calculation is not concerned with the relative position of a track within the greater traffic; it is concerned with how the relative position of that track within the traffic changes over the duration of the path. This calculation is first performed by taking the mean of the z-values for a given track, $\overline{x}_z^t$, where t denotes the track number. The sample

standard deviation of the z-values for each track, s_s^t, is then calculated with Eq. (46) where N_t is the number of test points. The distribution of the random variable X_S is based on the s_s^t from all tracks.

$$s_s^t = \sqrt{\frac{1}{N_t - 1}\sum_{i=1}^{N_t}\left(z_i^t - \overline{x}_z^t\right)^2} \tag{46}$$

Figure 50 shows the distribution of X_S for all of the tracks used in the Fort Carroll problem based on the 21 test points introduced above. Unlike the distribution of X_D^i, the distribution of X_S is approximated with a gamma distribution. The blue line in Figure 50 denotes the gamma distribution approximation of X_S with $a = 2.66$ and $b = 0.176$ as defined in Eqs. (47) and (48) [26].

$$X_S \approx f(x|a,b) = \frac{1}{b^a \Gamma(a)} x^{a-1} e^{\frac{-x}{b}}, \qquad x \geq 0 \tag{47}$$

$$\Gamma(n) = \int_0^\infty e^{-x} x^{n-1}\, dx\,, \qquad n > 0 \tag{48}$$

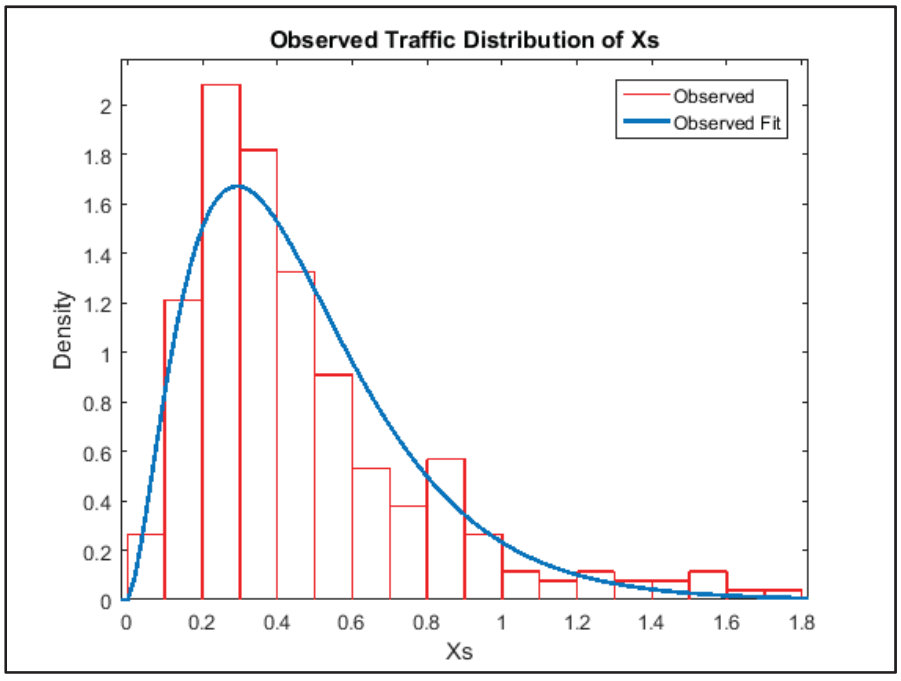

Figure 50: Distribution of X_S for observed traffic

The distribution of X_S for the ISR traffic is shown in Figure 51. The ISR X_S is also approximated with a gamma distribution with $a = 3.78$ and $b = 0.159$. As the ISR traffic distribution is only based on a generated set of 21 tracks, the ISR traffic is not as successfully approximated with a gamma distribution as the observed traffic that is based on over 250 tracks.

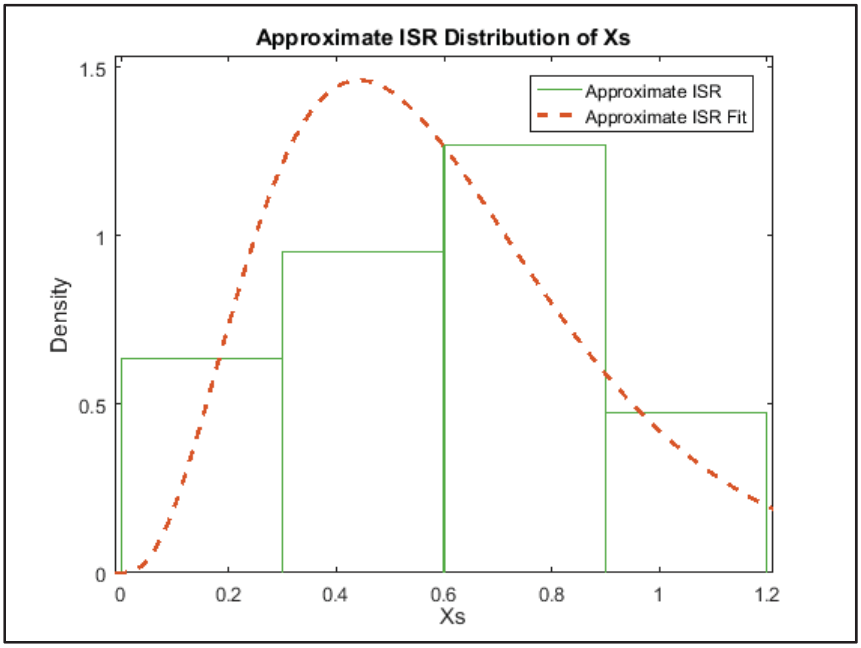

Figure 51: Distribution of X_S for ISR traffic

The portion of the anomaly detection process that depends on X_S takes the form of hypothesis testing in a similar form to that which is performed on X_D^i. Similar one sided null and alternative hypotheses are made at the significance levels found in Table 3. However, because X_S models the behavior of the tracks for the entirety of their lengths, the three hypothesis tests, one at each security state, only need to be conducted once. Table 3 indicates that the detection rates with X_S tend to be significantly less than they are with X_D^i at the same security state. This suggests that the behaviors of the ISR tracks are not as easily distinguished from the normal maritime tracks when X_S is used compared with X_D^i in this mission example. Although this indicates that X_S is not as effective of an identification parameter as X_D^i, it also indicates that the ISR paths developed in the path planning process succeed in blending in with the normal vessel traffic. The process is successfully able to plan paths that allow the USV to have a low identifiability during its mission.

Table 3: Adversary security states for X_S

Security State	False Alarm Rate (α)	Type II error	Detection Rate
Relaxed Security	1%	97.9%	2.1%
Normal Security	5%	89.7%	10.3%
Heightened Security	10%	80.1%	19.9%

Figure 52 shows the observed and ISR traffic PDFs and the location of the gamma values corresponding with each security state.

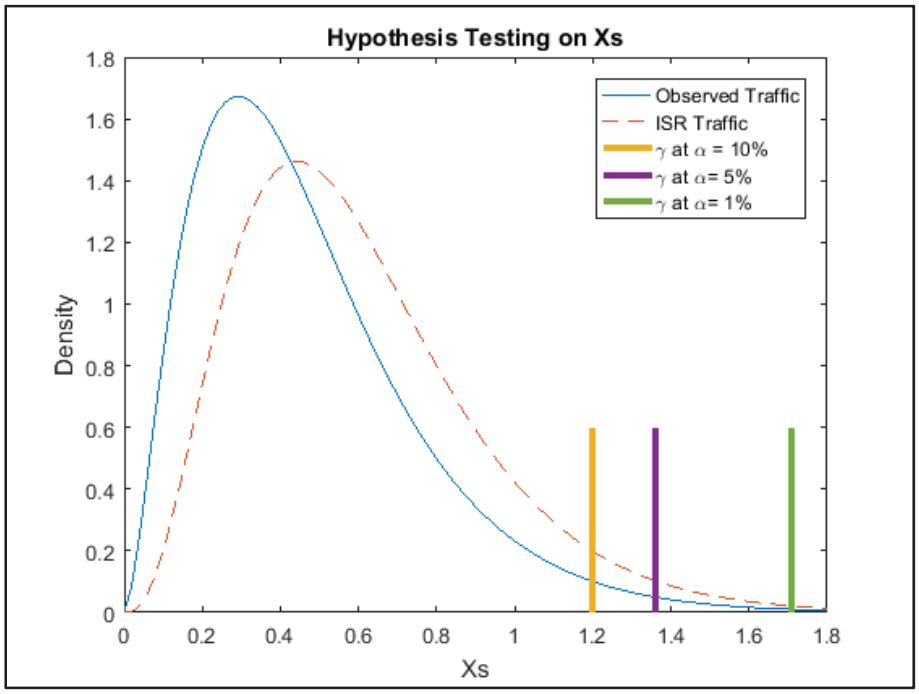

Figure 52: Hypothesis testing of X_S

Final Evaluation

With both the quality of intelligence and risk of identification frameworks established, it is possible to fully evaluate potential USV paths for an ISR mission. The intent of this path planning process is not to generate a "one-fits-all" path for a USV to follow for a particular target. The objective is to provide the mission commander with a range of paths, conservative to aggressive, that can be used based on the specific circumstances of the mission. Table 4 shows the evaluation of 21 generated optimal paths for the Fort Carroll example problem. These are the same 21 optimal paths used to generate the approximate ISR distributions of X_D^i and X_S. Each path is evaluated with a total of 66 hypothesis tests. 22 hypothesis tests are conducted at each of the three simulated security states. Of these 22 hypothesis tests, 21 of them are the hypothesis tests conducted on X_D^i at each of the 21 test points, and the final hypothesis test is the one conducted on X_S. Table 4 shows the number of hypothesis tests on X_D^i at each security state where the null hypothesis was rejected. None of the paths failed any of the hypothesis tests conducted on X_S. The table also shows the average distance, $\overline{D}_{i,k}$, and the percentage of observed tracks of which the USV's path is closer to the target.

Table 4: Fort Carroll path set evaluation

Path Number	Weight W	# of X_D^i Tests Failed $\alpha = 10\%$	# of X_D^i Tests Failed $\alpha = 5\%$	# of X_D^i Tests Failed $\alpha = 1\%$	X_S Hypothesis Tests Failed	Highest Allowed Security State	Average Distance (meters) $\overline{D}_{i,k}$	Average Distance Percentile
1	0.00	0	0	0	None	Heightened	518.5	54.5%
2	0.05	0	0	0	None	Heightened	503.4	64.5%
3	0.10	0	0	0	None	Heightened	489.0	75.6%
4	0.15	0	0	0	None	Heightened	475.8	80.7%
5	0.20	0	0	0	None	Heightened	463.3	87.8%
6	0.25	9	0	0	None	Normal	451.6	91.4%
7	0.30	16	1	0	None	Relaxed	440.5	94.6%
8	0.35	17	5	0	None	Relaxed	429.9	95.3%
9	0.40	17	13	0	None	Relaxed	420.4	98.5%
10	0.45	17	16	0	None	Relaxed	410.8	98.9%
11	0.50	18	16	3	None	Minimal	402.2	98.9%
12	0.55	18	16	5	None	Minimal	393.4	99.6%
13	0.60	18	16	7	None	Minimal	385.6	100%
14	0.65	18	17	12	None	Minimal	377.8	100%
15	0.70	18	17	12	None	Minimal	370.3	100%
16	0.75	18	17	12	None	Minimal	364.5	100%
17	0.80	18	18	13	None	Minimal	357.1	100%
18	0.85	18	18	15	None	Minimal	350.0	100%
19	0.90	17	17	14	None	Minimal	344.4	100%
20	0.95	17	17	14	None	Minimal	338.5	100%
21	1.00	17	17	14	None	Minimal	330.6	100%

The path evaluation results in Table 4 reveal that the generated USV paths perform as expected in regards to the balance between the quality of intelligence and the risk of identification. Conservative paths, those with low weights, lead the USV to collect lower quality intelligence but have a lower risk of identification. Aggressive paths, those with higher weights, are able to collect better intelligence but also come with a higher risk of identification. The seventh column of Table 4 indicates the highest security state of the adversary in which the USV is able to follow the given path and avoid being identified. More conservative paths allow the USV to operate without being identified even when the adversary is in a heightened security state. On the other hand, aggressive paths are likely to result in the USV being identified even if the adversary is in a relaxed security state. The most aggressive paths, those numbered 11-21, are only able to be followed by the USV if the adversary's security forces are minimal or nonexistent.

The best path options to present to the mission commander are the paths with the highest quality of intelligence at each adversarial security state. Paths 5, 6, and 10 are highlighted in Table 4 to indicate that each of these paths provide the best intelligence for a given security state. For example, if the mission commander expects that the adversary is in a heightened security posture, then Path 5 is the best route that the USV can follow and still avoid being identified by the adversary. Figure 53 shows Paths 5, 6, and 10 overlaid on the filtered vessel tracks used to generate the mean path. The inset on the right shows the three distinct paths and Fort Carroll. The path closest to Fort Carroll is Path 10, the middle is Path 6, and the furthest from Fort Carroll is Path 5. Figure 54 shows the same three USV solution paths overlaid on all of the observed vessel tracks in the Fort Carroll area.

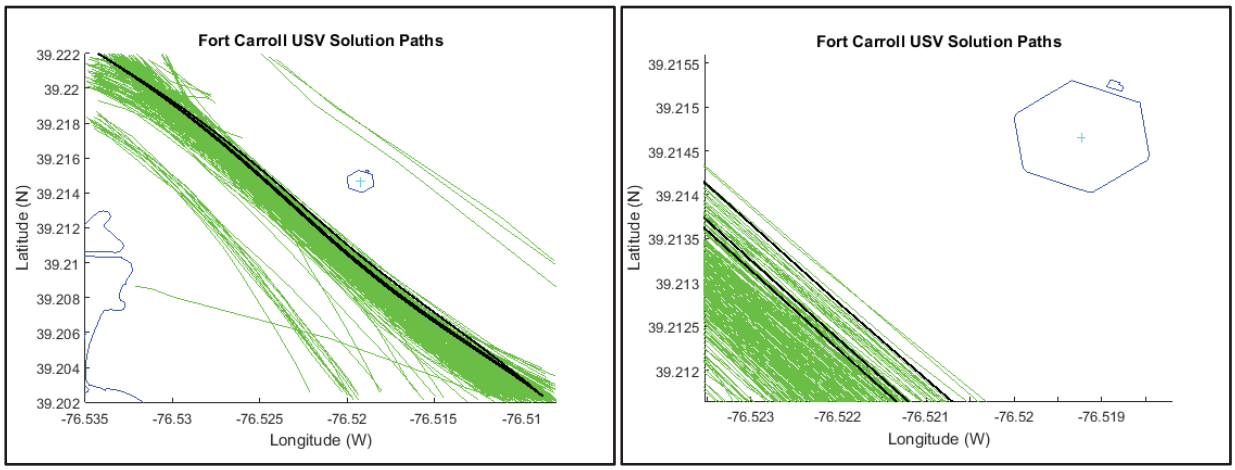

Figure 53: Three USV solution paths

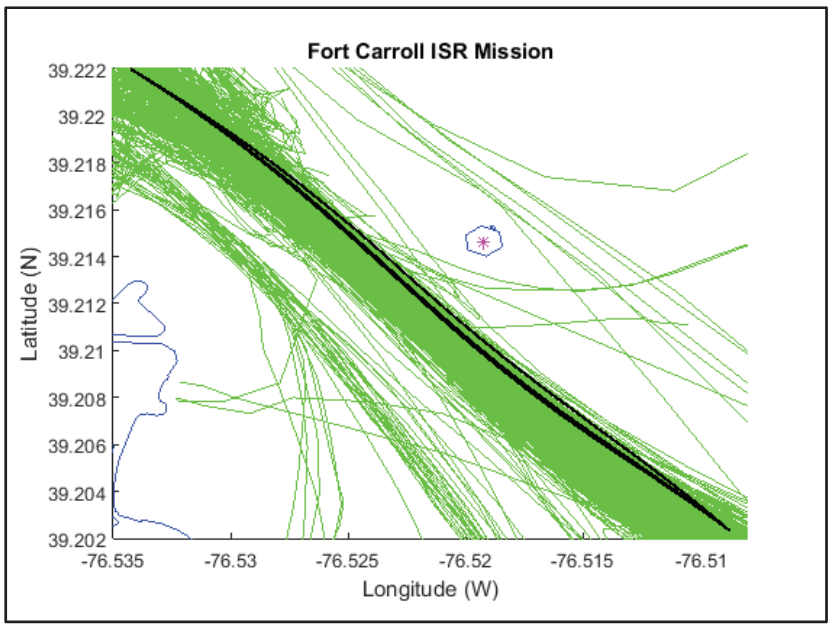

Figure 54: USV solution paths and all observed traffic

VIII. Method Validation using Independent Data Sets

The path planning process is developed on the Fort Carroll example problem discussed at length above. This problem is selected because of the simple traffic patterns surrounding Fort Carroll. Once the path planning process has been developed, it is important to test the process on targets other than Fort Carroll and with different sets of vessel tracks. This validation step helps to identify any areas of the process that have been biased by the unique characteristics of the traffic data or the Fort Carroll target. The objective of this research is to develop a path planning process that is flexible and applicable to a wide range of ISR missions. The process is validated with two example problems, one on Cape May at the entrance to the Delaware Bay and one on Point No Point in the Chesapeake Bay.

Cape May Mission

Cape May, New Jersey is located at the northern end of the entrance to the Delaware Bay. The Cape May lighthouse at the tip of the cape is considered to be the target for the ISR mission. Figure 55 shows the nautical chart of the area and Figure 56 shows the coastline in MATLAB [11, 27].

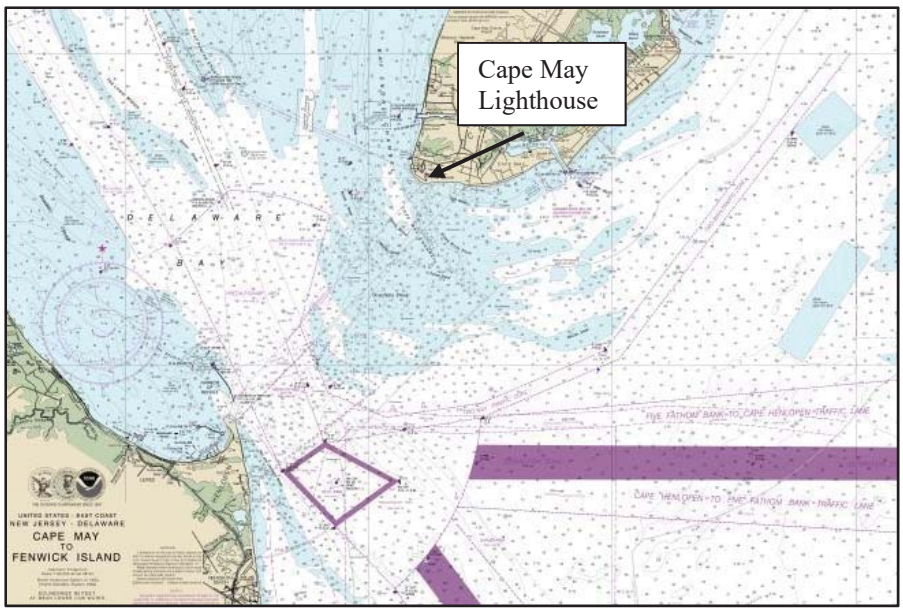

Figure 55: Nautical chart of Cape May **[27]**

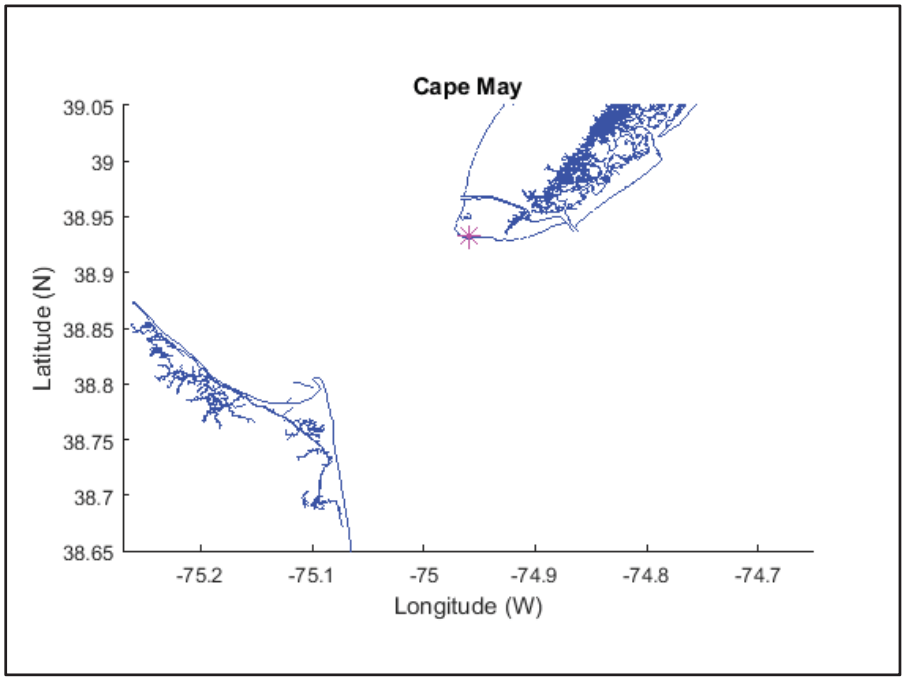

Figure 56: Cape May shoreline

There is a significant amount of vessel traffic in the Cape May area. A set of 2,459 vessel tracks in the region is shown in Figure 57. All of these tracks are taken from the same July 2009 AIS data set used for the Fort Carroll problem. However, because Cape May and Fort Carroll are in separate geographic locations, the problems do not share any of the same vessel track data.

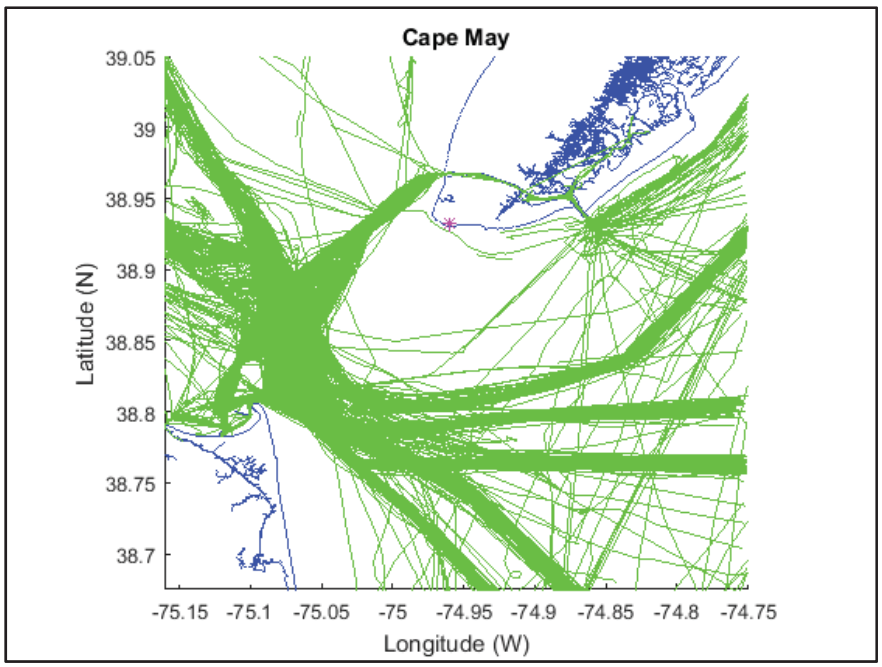

Figure 57: Observed vessel tracks near Cape May

The complex traffic patterns surrounding Cape May provide a number of options for the USV to follow in order to gather intelligence on the Cape May lighthouse. One of the most realistic mission profiles is for the USV to travel down the coast and up the Delaware Bay. Figure 58 shows the tracks of inbound vessels that travel down the New Jersey coast and up the Delaware Bay. The black line in the figure denotes the calculated mean path. Because the Cape May mission involves planning a longer and more complicated USV path, the mean path is generated in three sections. This allows the mean path to remain an accurate representation of the observed traffic for its entire length.

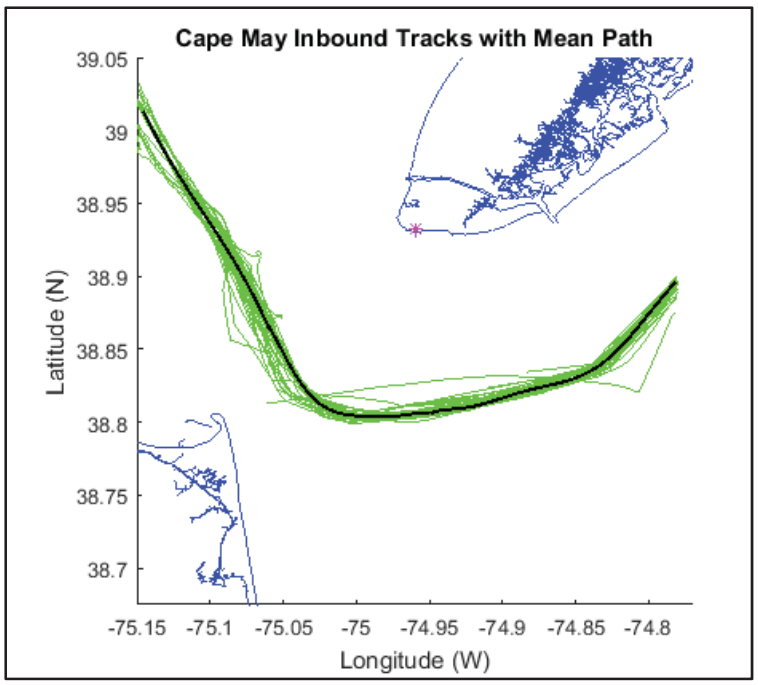

Figure 58: Selected Cape May tracks

Paths are generated at a wide range of weights, W, in order to determine the weights that generate usable paths for the USV. Because the geographic area covered in the Cape May mission is much greater than that found in the Fort Carroll mission, the Cape May weights must be much lower in order to keep the USV's path close to the mean path. Table 5 shows a summary of each of the best paths at each allowed security state. Path 1, which allows the USV to remain unidentified even if the adversary is in a heightened security state, allows the USV to draw closer to the target than 82.3% of the observed paths. Path 2, which is safe for the USV to follow if the adversary is in a normal security state, allows the USV to draw closer than 90.3% of the observed traffic. Path 3 allows the USV to come even closer but is only safe to follow under relaxed security conditions.

Table 5: Summary of Cape May mission path set evaluation

Path Number	Weight W	Highest Allowed Security State	Average Distance Percentile
1	0.010	Heightened	82.3%
2	0.015	Normal	90.3%
3	0.020	Relaxed	93.6%

Figure 59 shows the three USV solution paths listed in Table 5, shown in black, and the tracks used as the source of the mean path. The figure on the right provides an inset view of the three distinct paths. In this figure, the top path is Path 3, the middle is Path 2 and the bottom is Path 1. Figure 60 shows the three USV paths overlaid on all of the observed traffic in the Cape May area.

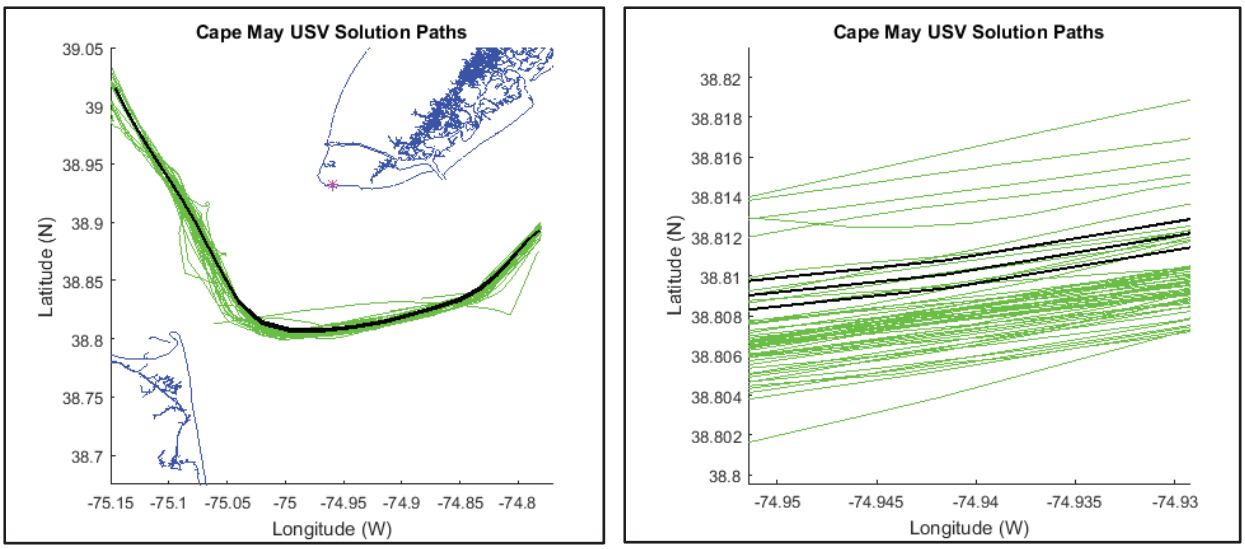

Figure 59: Three USV solution paths

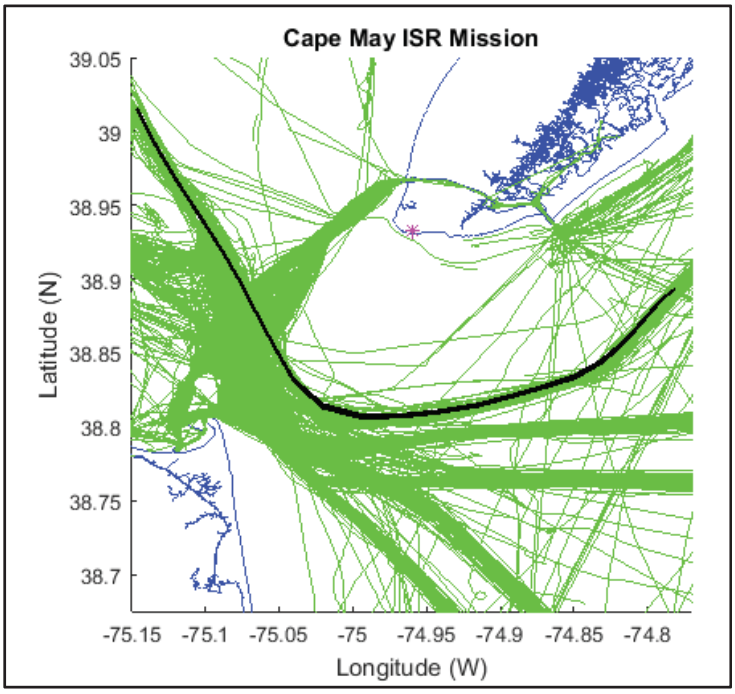

Figure 60: USV solution paths and all observed traffic

The path planning process is successfully able to develop paths for a USV to conduct ISR operations on Cape May. The Cape May example proves that the path planning process can be applied to complex traffic patterns and ISR missions covering larger geographic areas than that seen in the Fort Carroll problem. The Cape May mission did reveal the need to use smaller weights in the path planning cost function when an ISR mission involves a large geographic area.

Point No Point Mission

Point No Point in Dameron, Maryland is located about halfway up the Chesapeake Bay just north of the mouth of the Potomac River. The point of land itself serves as the simulated target for the ISR mission. The area of interest is shown in the nautical chart in Figure 61 and the coastline is shown in MATLAB in Figure 62 [11, 28].

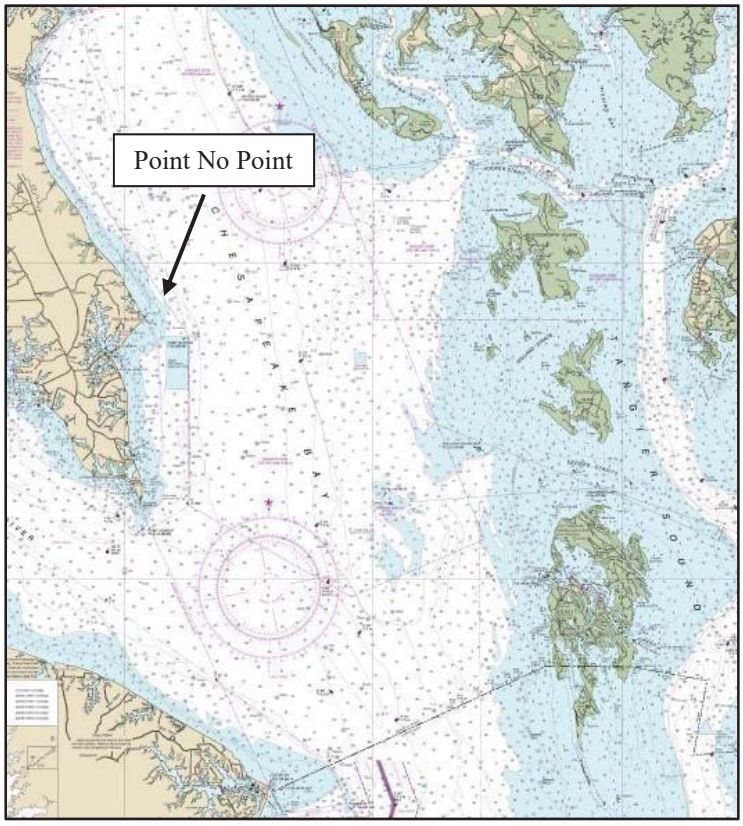

Figure 61: Nautical chart of Point No Point **[28]**

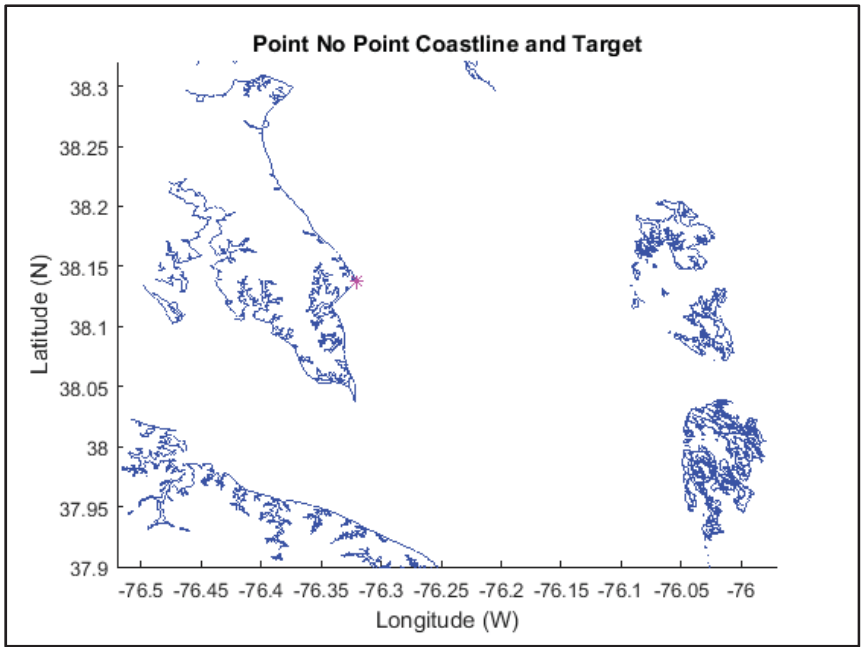

Figure 62: Point No Point coastline

The vessel traffic in the area near Point No Point is shown in Figure 63. This figure shows the 631 vessel tracks contained in the same July 2009 data set used for the other example problems. The majority of the vessels are travelling up and down the Chesapeake Bay. There are also tracks of vessels travelling into and out of the Potomac River on the left side of the figure.

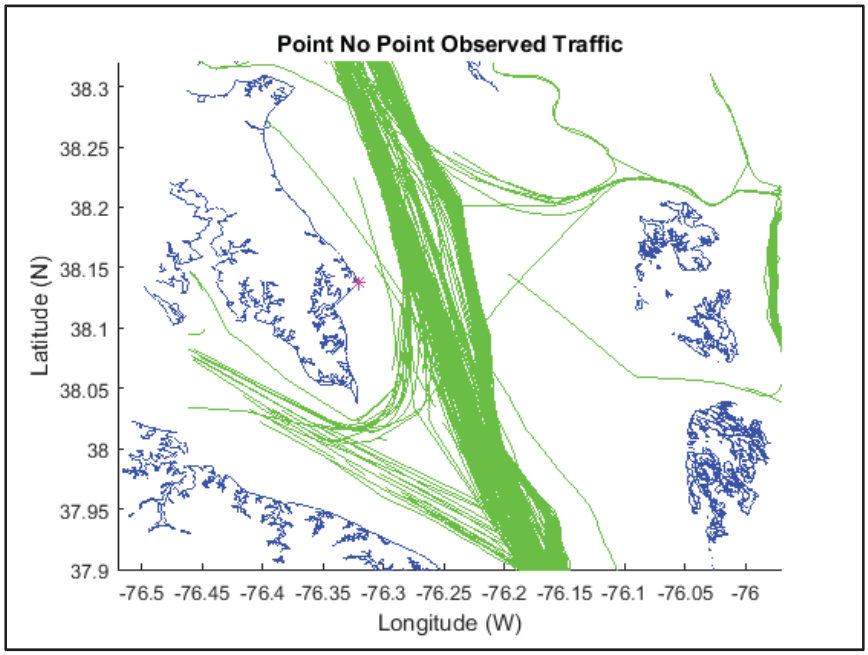

Figure 63: Point No Point observed traffic

The mission profile for the Point No Point mission is that of the USV travelling up the Chesapeake Bay from south to north and passing by Point No Point. Figure 64 shows the filtered tracks of vessels travelling north. The tracks have also been filtered to only contain the vessels that took the western route up the bay as this route brings the vessels closer to Point No Point.

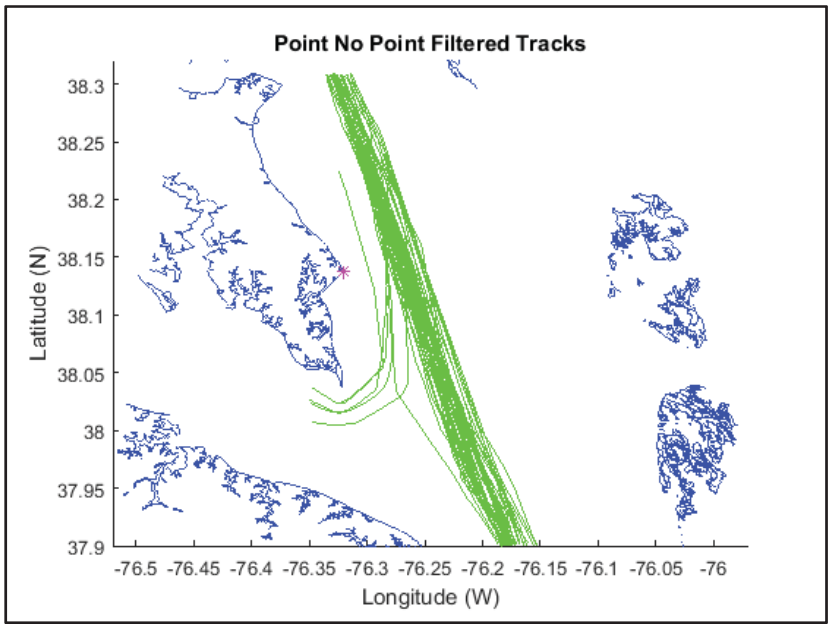

Figure 64: Point No Point filtered tracks

One of the challenges with the Point No Point problem is the fact that the vessel tracks are nearly vertical with respect to the latitude and longitude coordinate frame. This makes it difficult to perform polynomial regression as there is not a strong relationship between the independent variable, longitude, and the dependent variable, latitude. The vertical nature of the tracks also leads to difficulty with the cost function used in the optimal control formulation of the problem. The cost function measures the vertical distance between the solution nodes and the mean path. This calculation is not as effective at generating a quality path when the mean path is nearly vertical. To work around this issue, the latitude and longitude coordinates are switched in order to calculate the mean path and to solve the optimal control problem. Figure 65 shows the filtered tracks and the mean path with the latitude and longitude coordinates switched. The resulting USV solution path is then converted back to its proper orientation for evaluation.

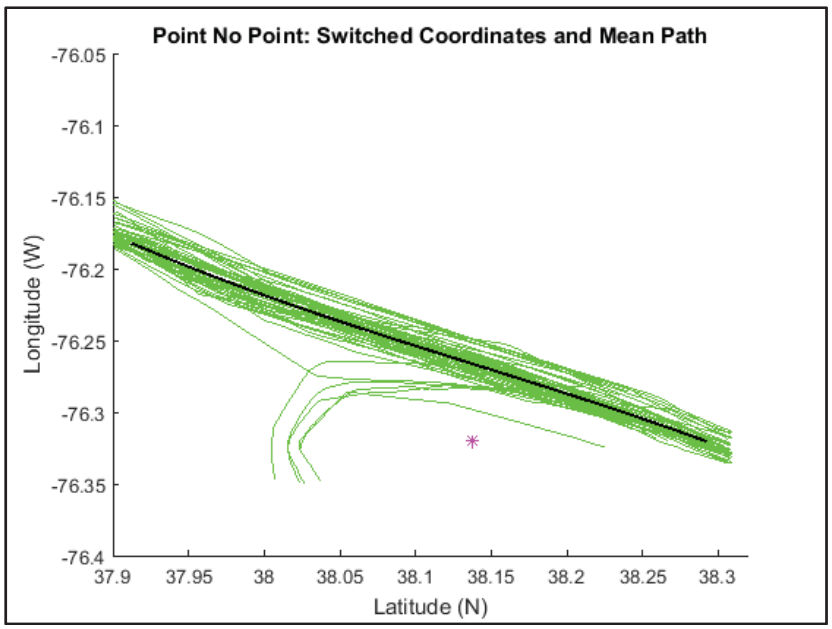

Figure 65: Switched coordinates and mean path

The USV optimal control problem is solved with a range of weights, W, in the cost function to develop a set of solution paths. Table 6 shows a summary of the best solution paths at each security state of the adversary. These results once again indicate that aggressive paths provide higher quality intelligence, but these paths should only be followed if the adversary is believed to be in a lower security state. The conservative Path 1 brings the USV closer to the target than 85.1% of the observed traffic while the aggressive Path 3 brings the USV closer than 97.0% of the observed traffic.

Table 6: Summary of Point No Point mission path set evaluation

Path Number	Weight W	Highest Allowed Security State	Average Distance Percentile
1	0.175	Heightened	85.1%
2	0.250	Normal	91.0%
3	0.500	Relaxed	97.0%

The three USV solution paths referenced in Table 6 are shown in Figures 66 and 67. Figure 66 shows the three paths overlaid on the filtered tracks used to calculate the mean path. The inset on the right shows Path 3 furthest to the west, Path 1 furthest to the east, and Path 2 in the middle. Figure 67 shows the three USV paths overlaid on all of the observed vessel tracks.

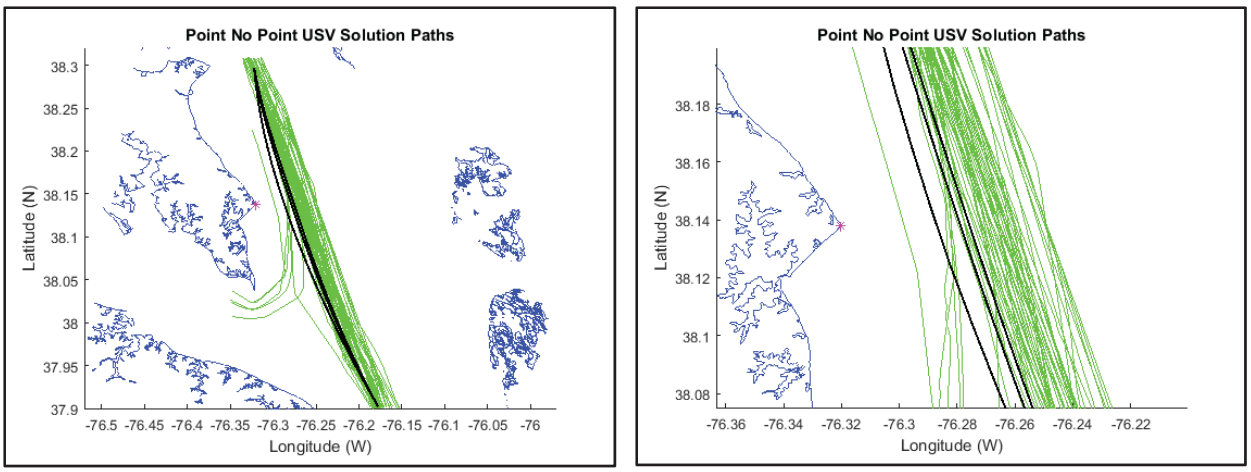

Figure 66: Three USV solution paths

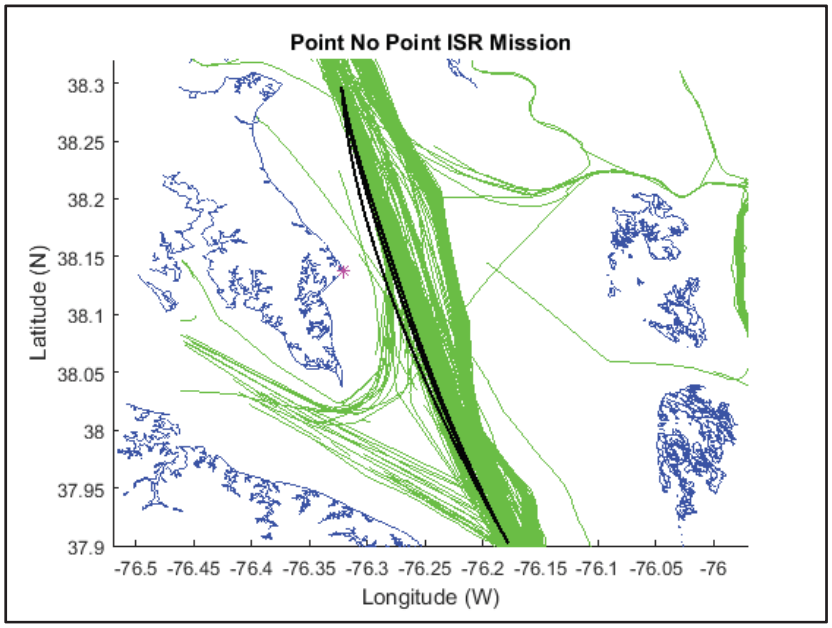

Figure 67: USV solution paths and all observed traffic

The path planning process developed in this research successfully generates a set of paths for a USV to follow when collecting intelligence on Point No Point. This indicates that the path planning process has the potential to be applied to a variety of mission scenarios. The Point No Point mission in particular demonstrates the path planning method's ability to deal with paths that are vertical or nearly vertical in relation to the latitude and longitude coordinate system.

The Cape May mission and Point No Point mission together demonstrate the flexibility of the developed path planning process. The Cape May mission demonstrates that the method can be used for problems involving large geographic areas by dividing the entire path into smaller sections. The Point No Point mission demonstrates the ability to switch the latitude and longitude

coordinates to solve paths that are close to vertical or are not able to be represented by a function in their original orientation. Both the piece-wise solution and coordinate switching can be used in tandem to develop paths for a wide range of ISR missions.

IX. Conclusion and Further Research

Conclusion

Unmanned Surface Vehicles promise to provide the Navy with the ability to conduct missions more effectively and in a manner that does not put Sailors in harm's way. Intelligence, Surveillance, and Reconnaissance of enemy targets is one area in which these vehicles can be of particular benefit to the service. This research set out to develop a path planning process to allow USVs to conduct ISR missions without being identified by data-based algorithms for behavior analysis. The developed path planning process is successfully able to go from observed vessel tracks to complete USV mission paths. The developed USV paths have a low risk of allowing the USV to be identified by the adversary while still allowing the vessel to collect quality intelligence.

Some portions of the path planning process are more mature than others. The path planning method has a strong data filtering and track sorting section. The Cape May and Point No Point missions also indicate that the mean path approach is effective at modeling the observed traffic. The path evaluation section is also strong in regards to hypothesis testing on X_D^i and the evaluation of the distance from the path to the target. However, the hypothesis testing on X_S is not an effective tool for the adversary to use to identify the ISR traffic. This is because the X_S distributions of the observed traffic and the ISR traffic are not distinct enough to make X_S an effective identification parameter. The speed selection and time in vicinity of the target sections are also not fully mature.

The result of this research is a path planning process that can be used for a variety of ISR missions. Although the focus has been on paths for USVs, these same paths can be used by manned ISR vessels. The process is flexible and robust enough to handle complex traffic patterns and a variety of targets. The process provides the mission commander with a set of solution paths that have a range of associated risks and benefits. The selected path is delivered to the USV as a baseline path to follow during its mission. The USV then deviates from this path as necessary to avoid other vessels and obstacles as it executes the mission.

Topics for Further Research

There are a number of areas in which this research can be expanded in order to develop a more sophisticated path planning process. Further research can be used to refine, streamline, and validate many steps of the process, but there are also areas where significant additions to the process can be made. One important topic for further research is modeling the spread of the observed vessel tracks around the mean path in the optimal control problem cost function. This calculation is important because the observed vessel tracks are narrowly collected along some portions of the path and are more spread out along other portions. The cost function currently does not take into account this spread of the normal tracks around the mean path when

calculating the cost of the distance between the solution path and the mean path. Including this traffic characteristic in the optimal control problem framework would allow for a more accurate traffic model.

Another area worthy of additional research is adjusting the path planning process to identify the best means of conducting "round-trip" ISR missions or repeated ISR missions on a target. This research focuses on how to make one pass on a target. It is beneficial to determine how a USV can discreetly enter and exit a target area without drawing suspicion during either its entrance or departure. This is particularly important when conducting ISR on a target in a harbor where the USV must both enter and exit the harbor within a short time frame. A related topic is determining how the USV can conduct routine ISR missions on the same target without being identified by the adversary.

The path planning process can also be adapted for use with moving targets. This application would identify the best way to gather intelligence on a moving vessel while still blending into the normal traffic patterns. An enemy warship entering or exiting port could present a good target for the USV. However, the USV would need to be cautious in how it approaches the warship as to not draw suspicion from the adversary.

References

[1] S. Savtiz *et al*., "U.S. Navy employment options for Unmanned Surface Vehicles (USVs)," RAND Corp., Santa Monica, CA, 2013.

[2] "Unmanned systems integrated roadmap FY2013-2038," Dept. of Defense, Washington, D.C., Rep. 14-S-0553.

[3] "The Navy Unmanned Surface Vehicle (USV) master plan," Dept. of the Navy, Washington, D.C., 2007.

[4] T. Huntsberger *et al.*, "Control Architecture for Robotic Agent Command and Sensing," *NASA Tech Briefs*, vol. 32, no. 10, pp. 5, Oct. 2008.

[5] D. Smalley, "The Future is Now: Navy's Autonomous Swarmboats can Overwhelm Adversaries," Office of Naval Research, 5 Oct. 2014. Available: http://www.onr.navy.mil.

[6] J. Hsu, "U.S. Navy's Drone Boat Swarm Practices Harbor Defense," IEEE Spectrum, 19 Dec. 2016. Available: http://www.spectrum.ieee.org.

[7] S. LaGrone, "Updated: Chinese Seize U.S. Navy Unmanned Vehicle," U.S. Naval Institute News, 16 Dec. 2016. Available: https://www.news.usni.org.

[8] C. Payne, *Principles of Naval Weapon Systems*, 2nd ed. Annapolis: Naval Institute Press, 2010.

[9] Chart: 12281, National Oceanic and Atmospheric Administration, 01 May 2014. Available: http://www.charts.noaa.gov.

[10] "MATLAB Overview," The MathWorks, Inc. Available: https://www.mathworks.com/products/matlab.html

[11] NOAA Shoreline Data Explorer, National Geodetic Survey. Available: http://www.ngs.noaa.gov/NSDE.

[12] "Automatic Identification System Overview," U.S. Coast Guard Navigation Center, 30 Jul. 2014. Available: http://www.navcen.uscg.gov.

[13] Vessel Traffic Data, National Oceanic and Atmospheric Administration Office for Coastal Management, 2014. Available: http://www.marinecadastre.gov/ais.

[14] "About ArcGIS," Esri. Available: http://www.esri.com/arcgis.

[15] P. Azariadis, "On using density maps for the calculation of ship routes," *Evolving Systems.* Springer, May 2016.

[16] "Least-Squares Fitting", *Curve Fitting Toolbox: User's Guide,* The MathWorks, Inc., Natick, MA, 2016.

[17] D. E. Kirk, *Optimal Control Theory: An Introduction*. Englewood Cliffs, NJ: Prentice-Hall, 1970.

[18] I. M. Ross, "A Beginner's Guide to DIDO," Elissar Global, Monterey, CA, 2007.

[19] K. P. Bollino *et al*., "Pseudospectral optimal control: a clear road for autonomous intelligent path planning," in *AIAA 2007 Conference*, Rohnert Park, CA, 7-10 May 2007.

[20] I. M. Ross, *A Primer on Pontryagin's Principle in Optimal Control*, 2nd ed. San Francisco: Collegiate Publishers, 2015.

[21] M. A. Hurni, "An information-centric approach to autonomous trajectory planning utilizing optimal control techniques," Ph.D. dissertation, Dept. Mechanical Eng., Naval Postgraduate School, Monterey, CA, 2009.

[22] J. van Laere and M. Nilsson, "Evaluation of a workshop to capture knowledge from subject matter experts in maritime surveillance," in *2009 12th Conf. Information Fusion*, Seattle, WA, 6-9 Jul. 2009, pp. 171-178.

[23] R. O. Lane *et al*., "Maritime anomaly detection and threat assessment," in *2010 13th Conf. Information Fusion*, Edinburgh, UK, 26-29 Jul. 2010, pp. 1-8.
[24] R. Laxhammar *et al*., "Anomaly detection in sea traffic - a comparison of the Gaussian Mixture Model and the Kernel Density Estimator," in *2009 12th Conf. Information Fusion*, Seattle, WA, 6-9 Jul. 2009, pp. 756-763.
[25] G. Pallotta *et al*., "Vessel pattern knowledge discovery from AIS data: a framework for anomaly detection and route prediction," *Entropy*, vol. 15, no. 6, pp. 2218–2245, Jun. 2013.
[26] K. S. Shanmugan and A. M. Breipohl, *Random Signals*, New York: John Wiley and Sons, 1988.
[27] Chart: 12214, National Oceanic and Atmospheric Administration, 01 November 2010. Available: http://www.charts.noaa.gov.
[28] Chart: 12230, National Oceanic and Atmospheric Administration, 01 January 2017. Available: http://www.charts.noaa.gov.